

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

(without alignments)
9657.174 Million cell updates/sec

Perfect score: 2501

Sequence: 1 cygcacgagatttgcagtc.....aaaaaaaaaaaaaaaa 2501

Scoring table: IDENTITY_NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summariles

Database : GenEmbl : *

GenEmb1:*

- 1: gb_ba:*
- 2: gb_hhg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_com:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_pl:*
- 25: em_pi:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_hhg_hum:*
- 31: em_hhg_in:*
- 32: em_hhg_other:*
- 33: em_hhg_mus:*
- 34: em_hhg_pln:*
- 35: em_hhg_rod:*
- 36: em_hhg_virt:*
- 37: em_hhg_vrt:*
- 38: em_hhg_vrt:*
- 39: em_hhg_hum:*
- 40: em_hhg_mus:*
- 41: em_hhg_other:*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	ID	Description
1	2501	100.0	2501	6	AX461660 Sequence
2	926.2	37.0	2316	8	AF360278 Arabidops
3	926.2	37.0	2381	6	AX461662 Sequence
4	926.2	37.0	2381	8	ATRH012747 Arabidops
5	540.8	21.6	160704	8	AP0032338 Oryza sat
6	321.4	12.9	2503	9	HSMLHIT
7	319.8	12.8	2484	6	AR122159 Human DNA m
8	319.8	12.8	2484	6	AR122159 Sequence
9	319.8	12.8	2484	6	AR131133 Sequence
10	319.8	12.8	2484	6	AX214172 Sequence
11	319.8	12.8	2484	6	AX234588 Sequence
12	319.8	12.8	2530	9	HS072743 Human DNA m
13	315.2	12.6	2068	3	BC006850 Homo sapi
14	300.6	12.0	2548	10	AY069160 Drosophill
15	299	12.0	2283	10	BC021815 Mus muscu
16	291	11.6	2274	10	AF250844 Mus muscu
17	261.4	10.5	2385	6	AR131262 Sequence
18	249.8	10.0	132171	3	AC008370 Drosophill
19	249.8	10.0	157559	2	AC020344 Drosophill
20	249.8	10.0	226833	3	AE003838 Drosophill
21	247.2	9.9	2233	3	AF068257 Drosophill
22	212.8	8.5	3218	6	AX214167 Sequence
23	212.8	8.5	3218	6	AX234583 Sequence
24	212.8	8.5	3218	8	SCMLHIT
25	208	8.3	41200	8	SC8520X
26	185.6	7.4	891	11	CNS06KRLN
27	141.6	5.7	727	6	AB6098
28	141.6	5.7	727	6	AR155591 Sequence
29	141.6	5.7	727	6	E66116 Genome DNA
30	135.4	5.4	2664	3	AF346620 Trypanoso
31	133	5.3	1613	1	AE005550 Escherich
32	133	5.3	243184	1	AP002568 Escherich
33	128.6	5.1	2446	1	STYMWUL
34	128.6	5.1	20418	1	AE0088904 Salmonell
35	128	5.1	10151	1	AE013665 Yersinia
36	128	5.1	214050	1	AT414142 Yersinia
37	127	5.1	249050	1	AT627283 Salmonell
38	125	5.0	1848	1	AF440199 Escherich
39	125	5.0	2300	1	ECMWTUG
40	125	5.0	12088	1	AE000489 Escherich
41	125	4.9	338534	1	EC00W93 Human DNA m
42	123.6	4.9	13668	1	AE004907 Pseudomon
43	117.4	4.7	2308	9	BC005833 Homo sapi
44	117.4	4.7	2308	9	BC005866 Homo sapi
45	116.6	4.7	13571	1	AE004982 Halobacte

RESULT 1

LOCUS	AX461660	2501 bp	DNA	linear	PAT 08-JUL-2002
DEFINITION	Sequence 1 from Patent WO0224890.				

VERSION AX461660.1 GI:21726826

SOURCE	Oryza sativa
...	...

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE
1 Ehrhartoidae; Oryzeae; Oryza

TITLE	DATE	BY	REMARKS
Rice mill ortholog and uses thereof			

Oy	1741	GATGTTGGAAACCTGGAGATTGGCAGAACTAAACACTGAGTACTATAAAGAAAATGCTGAG	1800
Db	1741	GATGTTGGAAACCTGGAGATTGGCAGAACTAAACACTGAGTACTATAAAGAAAATGCTGAG	1800
Oy	1801	ATGATTAAATGAGTCCTTTCTTATTCACATTGTATCAAGAATGGCAAAATTGACAAGACTTCT	1860
Db	1801	ATGATTAAATGAGTCCTTTCTTATTCACATTGTATCAAGAATGGCAAAATTGACAAGACTTCT	1860
Oy	1861	GTTGTAACAGCCAGTACACCCTCGATATATGCAACGCTCTTCCAGAAATTTGTGTGGCTTTA	1920
Db	1861	GTTGTAACAGCCAGTACACCCTCGATATATGCAACGCTCTTCCAGAAATTTGTGTGGCTTTA	1920
Oy	1921	GGAATGATGTACTTGGGATGAGCAGAGAAAGAGTCTTCAGAACGTAAGCTTCGCTCTGA	1980
Db	1921	GGAATGATGTACTTGGGATGAGCAGAGAAAGAGTCTTCAGAACGTAAGCTTCGCTCTGA	1980
Oy	1981	GGAAACTCTTATGCACCTCATCCGCCAATCCCTCCMAATCCATCTCGGGAATGGCAATTCAT	2040
Db	1981	GGAAACTCTTATGCACCTCATCCGCCAATCCCTCCMAATCCATCTCGGGAATGGCAATTCAT	2040
Oy	2041	TTTATCAAGAAAAATAGAGATTCAATGGCTGATGAACATGCTGAGATGATCTAATATCA	2100
Db	2041	TTTATCAAGAAAAATAGAGATTCAATGGCTGATGAACATGCTGAGATGATCTAATATCA	2100
Oy	2101	GATGAAATGACGTTGATCAAGAACTCTTGGCGGAAGAGAAAGCAATGGGCCCAAGT	2160
Db	2101	GATGAAATGACGTTGATCAAGAACTCTTGGCGGAAGAGAAAGCAATGGGCCCAAGT	2160
Oy	2161	GAGTGGACCATTTACGATGCTGTTTTCTTCATTCATGACGACTTTTCTCAAGCCCCGAAG	2220
Db	2161	GAGTGGACCATTTACGATGCTGTTTTCTTCATTCATGACGACTTTTCTCAAGCCCCGAAG	2220
Oy	2221	TCAAAGGCAACAGATGAGAAAGTTTGTGCAAGTGTGCTTCTTGGAGAAAATCTCACAAAGT	2280
Db	2221	TCAAAGGCAACAGATGAGAAAGTTTGTGCAAGTGTGCTTCTTGGAGAAAATCTCACAAAGT	2280
Oy	2281	TTTGAAGAAGTGTAGCTATATAAGTAGAGAAATGAAGGAGAGATAGATCATGATCATAGG	2340
Db	2281	TTTGAAGAAGTGTAGCTATATAAGTAGAGAAATGAAGGAGAGATAGATCATGATCATAGG	2340
Oy	2341	AGTGTTTTGAAGATGTGTATATATTTCAACCGTATTATGACTTGTGATAGTGTCTGTAGAA	2400
Db	2341	AGTGTTTTGAAGATGTGTATATATTTCAACCGTATTATGACTTGTGATAGTGTCTGTAGAA	2400
Oy	2401	ACTGAAGAAAGAAAGATGGCTTACTTGTGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGT	2460
Db	2401	ACTGAAGAAAGAAAGATGGCTTACTTGTGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGT	2460
Oy	2461	TATTCTGATCAACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2501	
Db	2461	TATTCTGATCAACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2501	

RESULT 2
AF360278

LOCUS
DEFINITION Arabidopsis thaliana putative MHL1 protein (Atg905140) mRNA,
complete cds.

ACCESSION
VERSION AF360278.1 GI:13430731

KEYWORDS
SOURCE FTL CDNA.

ORGANISM Arabidopsis thaliana.

REFERENCE
AUTHORS Arabidopsis thaliana.
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 2316)
Yamada,K., Iku,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
Jones,T., Kamuya,A., Karlin-Neumann,G., Kawaji,J., Kim,C., Lam,B.,
Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,

TITLE	Satou, R.M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
JOURNAL	Arabidopsis Full Length cDNA Clones
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 2316)
TITLE	Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Chung, M.K., Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
JOURNAL	Submitted
COMMENT	Direct Submission Submitted (13-MAR-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.
FEATURES	source
source	The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Chung, M.K., Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Lam, B., Lin, J., Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Davis, R.W., Ecker, J.R. and Theologis, A.
gene	Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.
CDS	Location/Qualifiers 1..2316 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /chromosome="4" /clone="RAFL09-15-M13 (R09268)" /note="This clone is in a modified p Bluescript vector (FIC-1) as a BamHI/XhoI insert. ecotype: Columbia" 1..2316 /gene="At4g09140" 1..16 /gene="At4g09140" 17..2200 /gene="At4g09140" /codon_start=1 /evidence=experimental /product="putative MHL protein" /protein_id="AAK25968.1" /db_xref="GI:13410732" /translation="MEEEESPATIVREPKIQRLEESVYVNRAGEVIORPVSAYK ELVENSIDADSSISVVKKGGLKLDVSDGHEITREDLPILCEHTPTSKLKEFDL FLSLSMEFREGELASMTVAHVYTTITDKQIHGRVSYRGVWEHPKCAAVKGTQ IMENLEFYENIARLKTLONSADYGGKILVDSRMAIYNNVSEFGRHGAVKADHVSU VSPRSLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT NDRIVECSALKRAIEIYVAATIPKASPEYVMSINPREHDIINHPKTEVSLNDB IITIMDISEVYKLRNANDRTREBOQVEYTIQSLTJSOKSDSPVSQKPSGQTKTVEV NKWVRDSDSPGRLHAFLOPKQSLPDKXSSLSVSSVQRNRPEETADLSVOEL IAGVDSCHPMGLMETVENCYVGMADVFLAYVQNIHLYANVNLSEKLAYQOTLRR EYFHNALQISDPALSELILALKEEDLDGNDKDLKERIAEMNTLEKEMKEE EYFPRLSISVSGVSAKNIMKVEVSSCGSPMEGPIISNVAKTLIVLFT

Query Match 37.0% Score 926.2; DB 8; Length 2316;
 Best Local Similarity 65.4%; Pred. No. 2,2e-178;
 Matches 1408; Conservative 0; Mismatches 733; Indels 12; Gaps 3;

158 GGGAGCCGCCCCGATCCGAGGTTGAGAGTCGGTGTGTAACCCGCGCGCGGG 217
 57 GAGAGCCAGCCAGATTCAGCGCTTAGAAGATAGTAGTACACCGTATCCAGCTGTG 116
 218 AGGTATATCAGCGCGCTGTGCGGCGGAGAGGCTCATGAGAAAGCGCTGCGCGTG 277
 117 AAGTATATCAGCGCTGTGCGGCGGAGAGGCTGTGAGAGAGCGCTGCGCGCG 176
 278 GCGCGTCCAGCGCTGTGCGGCGGAGAGCGGTGCGCTCAAGCTCAATCCAGCTCCG 337
 177 ATTCAGATTCATTAAGCGCTGTGCAAGAGCGGTGTTGAATCATTCAGCTCCG 236
 338 ATGAGCGCCATGCGATCAGGTTGAGATTTGGCAATTTGTGCAAAAGCGCTACCT 397
 237 ACAGCGGTACGCGTATAGAGCTGAAGACTTCCGATCTACTATGCGAGAGCATACAT 296
 398 CAAAGTATATGATACAGAGATTCGAGACCATTAATCATTTGGGGTTGAGGGGAG 457
 297 CGAAGCTGATAGTTGAGATTTGCTCTGTAGTTCAATGGGATTTAGAGAGAG 356
 458 CTTGGCTAGTATGATATGTTGGCATGTTACCGTACAGATTAACAGAGCGCAT 517
 357 CATTAAGTATGATACGCTATGCTGCTCATGTTACAGTACATCTACTTAAGAGCGCA 416
 518 TGCAGCGCTACAGGCTTTCTTACAGAGATGTTGATGAGAGATGAGCTTAAGCTTGG 577
 417 TTCAGGTATATAGAGTGTCTTATAGAGATGTTGATGAGATGAGATACCAAGGGGTGG 476
 578 CTGGCGTGAAGAGACTGAGTATGTTGAAATCTATTTTACAACTGAGAGCGCA 637
 477 CTGGCTGAAGAGAGATGAGTATGTTGAAATCTATTTTACAACTGAGAGCGCA 536
 537 GGAAGAGCTTCAAAATTTGCTGATGATTTACGGGAAATCTGATTTGCTGAGCGGA 596
 638 AGAAGAGCTTCAAAATTTGCTGATGATTTACCGGAAATCTGATTTGCTGAGCGGA 697
 537 GGAAGAGCTTCAAAATTTGCTGATGATTTACGGGAAATCTGATTTGCTGAGCGGA 596
 698 TTGAGTCCATCATCAACGTTACTTCTGCGAAGAGAGTGGAGCATATAGAGCA 757
 597 TGGCTATTTATCAATTAATGTCAGCTTTCTGCGAAGAGAGTGGAGCATATAGAGCA 656
 758 ATGTCATATGTCAGATACATCTCCTCAAGGTTAGATCTATCAAGAGTGTCTATGGGCT 817
 657 ATGTCATATGTCAGATACATCTCCTCAAGGTTAGATCTATCAAGAGTGTCTATGGGCT 716
 818 CTGCTGCTGATCTCAATGAAGATTAAGGTTTCAATATGAGATGCTGAGATTCATCT 877
 717 CAGTTGCAAGAGCTTGAAGAGTGAAGTTCCTGCTGAGTCTGCTGCTGCTGCTGCT 776
 878 TCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 937
 777 TGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 836
 938 TTTTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 997
 837 TTTTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 896
 998 ACTCTGCAACATTTGCTCAAGCATCAACCTTTCAATATGATGATGATGATGATGATG 1057
 897 ATGCTGCAACATTTGCTCAAGCATCAACCTTTCAATATGATGATGATGATGATGATG 956
 1058 CAGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1117
 957 GGAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1016
 1118 AGCGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1177
 1017 AATATCATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1076
 1178 CAGAGATATTCAGAGCTGAGCATTAACCTTATCAGAGATTTGCTCAAGCTAACCCACA 1237

1077 CTAGAGAGCTTCAAGAGCAGAGAGTGAATATCATTCATTCATTCATTCATTCATTCAT 1136
 1238 AGATTAAGTGTTCAGAGCAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1297
 1137 GTGAT---TCTCAGTGTTCAGAGCAGAGCTTCTGACCAAAAGAGAGAGAGAGAGAG 1193
 1298 GCCAATATGTCAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAG 1357
 1194 ACAAAATGTCAGAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAG 1253
 1358 GGCATCTTCAATCT---TGAAGAGAAATTTGATCTGATGATGATGATGATGATGATG 1414
 1254 CCAAGCCAGAGAGCTTCTGACAGAGGTTCTAGTGTGATGATGATGATGATGATGATG 1313
 1415 GATCAG 1474
 1314 GCGAAG 1373
 1475 AATATGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGAT 1534
 1374 GAGTGAAG 1433
 1535 GACTTCCAGATGAG 1594
 1434 GAATGAG 1493
 1595 TGTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1654
 1494 TGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1553
 1655 ATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1714
 1554 AGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1613
 1715 AGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1768
 1614 AGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1673
 1769 TAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1828
 1674 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1733
 1829 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1888
 1734 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1793
 1889 TGGACGCTGTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1948
 1794 TGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1853
 1949 AAGATGCTTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2008
 1854 AAGATGCTTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1913
 2009 TCCCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2068
 1914 TTTTCCCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1973
 2069 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2128
 1974 AGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2033
 2129 TTGCGAAG 2188
 2034 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2093
 2189 CATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2248
 2094 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2153
 2249 AGTGTCTTCTGAG 2301

Db 2154 AGGTACGATCCCTTGAAAGCTGTACAGATATTGCAAGCATGCTAATGAA 2206

RESULT 3
AX461662 2381 bp DNA linear PAT 08-JUL-2002
LOCUS Sequence 3 from Patent WO0224890.
DEFINITION AX461662
ACCESSION AX461662
VERSION AX461662.1 GI:21726828
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE
1 Mahajan, P. B.
Rice mhl ortholog and uses thereof
Patent: WO 0224890-A 3 28-MAR-2002;
PIONEER HI-BRED INTERNATIONAL, INC. (US)

FEATURES
source
1. 2381
Location/Qualifiers
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
3. 2216
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD38486.1"
/db_xref="GI:21726829"
/translation="MIDSSILFAEMEEESPATTVIPREPKIORLEESVNRVIAAG
VTORVSAVKELVENSIDSSISVYVVDGKLQVSDHGIREDLPILCEHNT
TSKLPKEDFELSWSGRFEBALSWTVAHVTTITKQIHGRVSRDVMEREP
RACAAVKGQIWMENLFYMIARRTIOLASADYKIDVLIRMIHYNVEFNSKH
GAVKADVHSVSPRLDSIRSYGVSAKNLKVESGDSGCEFMDEGTSNRY
AKRTILVEINRLVESCALKRAIEIVYATLPKSPVYSINLPREHVDINHT
KEVELINDELIIEMIOSFEVKLRNANDTFROKXVYIOTSLTSOKSDSPVQKP
SGOKRQPVNKMVRDSDSPGRILHAFIQPRPOSIPKVSLSVRSVRSRPNKE
TADLSSVDELINAGVDSCHPGLVETRNCTVGMADVFAIVQYNTHLIVANVNSK
ELMYOOLRFHFAPNLIQSDAPLSELILALIKEDLDPGNDIKDLIERLANVNE
LKKEAEMLEEFYSIHDSANILSRPLVLDQYTPDMVDPPEFLICIGDVEDEKS
CEQVSAIAGNFYAHMPLPNSGDIQYFYSKRGSSQEKSDLEBNVMEINLDDLL
LSDANAMQRMSTIOHVLFPSPMLFELKRPASMAISNGTVKVASLEKLYKIERC"

BASE COUNT 721 a 450 c 548 g 662 t

ORIGIN

Query Match 37.0%; Score 926.2; DB 6; Length 2381;
Best Local Similarity 65.4%; Pred. No. 2.2e-178;
Matches 1408; Conservative 0; Mismatches 733; Indels 12; Gaps 3;

QY 158 GGGAGCCGCCCGCCGATCCGAGGTGAGAGATGCGGTGTAACCCGATCGCGGGGG 217
Db 73 GAGAGCCACCGAAGATTCACAGCTTAGAAGAAATCAGTAGCAACCGATCGAGCTGTG 132
QY 218 AGGTATTCACAGCGCGGTGCGGGGTGAAGAGCTCATCAGGAACAGCTGAGCGCTG 277
Db 133 AAGTAATCCAGCGCTCAAGTTTCAGCTGTGAAGAGCTCGTTGGAACAGCTGAGCGCG 192
QY 278 GCGCCTCCAGCGCTCCGCTGCGGTGAAGAGCGTGGCCCTCAAGCTCATCCAGGTCTCG 337
Db 193 ATTCAGTTTCCATTAAGCGTGTGTCAAAGACGGTGTGGAACATCATCAAGTCTCCG 252
QY 338 ATACAGCGCATGCGATCAGTTTGAAGATTTGGCATATTGTGCGAAGAGCATACTACT 397
Db 253 ACAGCAGCTCAGGTATTAGACGTGAAGACTTGCAGTACTATCGAGAGCATACACAT 312
QY 398 CAAGATTATGATACAGAGATCTGCAGACCAATTAATGATGGGGTGTAGAGGGAGG 457
Db 313 CGAAGCTGAAGATTGAGAGATTGTCTCTGAGTTCAATGAGGATTTAGAGGAGAGG 372
QY 458 CTTTGGCTAGTATGACTTATGTTGGCATGTACCGTGACAGACGATTAACAGAGCCAT 517
Db 373 CATTAGCTATGATGACTATGTTGCTCATCTTACAGTACTACTACTTAAGGCCAGA 432

QY 518 TGCAGCGCTACAGCGTTTCTTACAGAGATGCTGTAATGAGAAATGACCTGCTCG 577
Db 433 TTCTAGTTTATAGAGTTCTTATAGAGATGCTGATGAGAGATGAAACCAAGCGCTG 492
QY 578 CTGCGGGAAGAGCAACATGATGTTGAAAATCATTTTCAACATGATGAGCCGCA 637
Db 493 CTGCTGCAAGAGACACATATATGTTGATGTTGTTCTCAATATGATGCTGAGAA 552
QY 638 AGAAACATTCAGAACTCCAAATGATGATACCCCAAGATGCTGATGCTGATGCTG 697
Db 553 GGAAGACACTTCAAAATTCGCTGATGATATACGGGAAATTCGAGATTTCGACCGCA 612
QY 698 TTGAGTCCATTCATCATTACAGCTTCTCTTGGCAAGATGAGCAATAGACAG 757
Db 613 TGCTATTCATTCATTAATTAAGTCAAGCTTCTCTGTCAGAAAGATGAGCTGTAAG 672
QY 758 ATGTTATAGTGCAGATACATCTCAAGGTAGATGATGATGATGATGATGATGATG 817
Db 673 ATGTTATAGTGCAGATACATCTCAAGGTAGATGATGATGATGATGATGATGATG 732
QY 818 CTGCTGCTGATCTCATAGAAATTAAGGTTTCATATGAGATGCTGCAATTCATCT 877
Db 733 CAGTTGCAAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 792
QY 878 TCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 937
Db 793 TTGATATGAGAGGTTTCTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 852
QY 938 TTTTCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 997
Db 853 TTTTCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 912
QY 998 ACTGTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1057
Db 913 ATGTTGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 972
QY 1058 CAGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1117
Db 973 GGGAACTGTTGATTCATTAATTCATTCATTCATTCATTCATTCATTCATTCATTC 1032
QY 1118 AGGTATTCAGAACTAAGAAATGCTTTGAGAAATGCTGATGATGATGATGATG 1177
Db 1033 AAATCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1092
QY 1178 CCAAGATATTCAGAACTCAGGATTAATTCATGAGGATGCTCAAGCTACCAACAA 1237
Db 1093 CTGAGAGCTTTCAGAGCAAAAGTGAATACATTCATTCATTCATTCATTCATTC 1152
QY 1238 AGATTAAGTTTCTGAGGCGATGAGGCTTCTGGAACAAATCTCAAAATTCCTG 1297
Db 1153 GTGAT--TCTCAGTTTCTCAGAGGCTTCTGGAACAAATTCCTGCTGTA 1209
QY 1298 GCCAAATGTCAGAGCAAGATTCAGCAATTCATGATGATGATGATGATGATGATG 1357
Db 1210 ACAAAATGTCAGAGCAAGATTCATGATGATGATGATGATGATGATGATGATGATG 1269
QY 1358 GCGAATCTTCAAACT--TGAAAGAAATTCATTCATGATGATGATGATGATGATG 1414
Db 1270 CCAAGCCACAAAGTCTCCGACAAAGTCTTGAATGATGATGATGATGATGATGATG 1329
QY 1415 GATCAGAGGAAACCAAAAGATGCTGATGATGATGATGATGATGATGATGATGATG 1474
Db 1330 GGCAGAAAGAAACCAAAAGATGCTGATGATGATGATGATGATGATGATGATGATG 1389
QY 1475 AAATAGATTCAGTTCATGCTGCGCTTGTGATGATGATGATGATGATGATGATG 1534
Db 1390 GAGTTGACACTGCTGATCAGATCAGATGATGATGATGATGATGATGATGATGATG 1449
QY 1535 GACTTGCAGATGAGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1594
Db 1450 GAATGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1509
QY 1595 TCGTAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1654

```

1
Db 1510 TGGGAACTCTAGCAAGAGTAAATGATACAGCAAACTTCGTCGTTTCTCATTTTA 1569
1565 ATGCTATTCAGCTCAGTGAACACAGCTCCATTCAGAGTTCCTGCTGATGACCTGAAG 1714
1570 ACGCAATACAGCTTACGATCCAGCCCTTTGTCAGATTTATTTGTTGGCTCTGAAG 1629
1715 ACGATGAAT-----GATGAGTGTGAAGAGATGATGAGAACTGAGATTCGAGAG 1768
1630 AGGAGAGCTCTGATCCAGCAATGATACAAAAGATGATCAAGAAAGAAATGCTGAAA 1689
1769 TAAACAGTGAATCTAAAGAAATGCTGATGATTAATGATCTTTCTTATTCACA 1828
1690 TGAATACAGAACTCTCAGAGAAAAGAGAAATTTAGAGAGATTTTACGCGTGACA 1749
1829 TTGATCAAGATGCAAAATTCACACAGACTTCTGTTGATCTGACCAAGTACACCTGATA 1888
1750 TTGACTCCAGTCAAAATTTGTCAGAGCTTCTGATGATCTGACCAAGTATACCTTGACA 1809
1889 TGGACCGCTCTCCAGAAATTTGTTGGCTTAAAGAAATGATGATCTTGGATGACGAGA 1948
1810 TGGATCGTGTCTGAAATTTTACTATGCTTGGGAAATGATGATGAGTGGAGATGAGA 1869
1949 AAGAGTGTCTCAGACAGTACCTTCTGCTGATGAGAACTCTATGACCTTATCCCCCA 2008
1870 AAGAGTGTCTTCAAGAGATTTCTGACAGTATGAGAACTTTTACGCAATGATCTCTC 1929
2009 TCCCTCCAAATCAGTCTGGGAAATGATGATTTATACAGAAATATGACATTAATG 2068
1930 TTTTCCCAAAACCCATCGGCTGAGGATTTCTGATGATGATGAGAGTGGAGCTCTC 1989
2069 CAGTGAACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2128
1990 AGGAAAGTCAATTTAGAGGATTAAGCTGATGATGATGATGATGATGATGATGATGATG 2049
2129 TTTCCGAGACAGAGCAGTGGGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2188
2050 TGTCAATGCTGAAAGAGATGAGTGGGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2109
2189 CATCATGAGATTTTCCCAAGCCCGGCAAGTCAAGTGGCAAGATGAGATGAGTGGTGG 2248
2110 CGTCAATGAGATTTTCCCAAGCCCGGCAAGTCAAGTGGCAAGATGAGATGAGTGGTGG 2169
2249 AGTGTCTCTCTGGAAGAACTCTACAGATTTTGAAGAGTGTGATGATGATGATGATG 2301
2170 AGGTACATCCCTTGAAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATG 2222

RESULT 4
ATH012747 2381 bp mRNA linear PLN 18-DEC-1999
LOCUS Arabidopsis thaliana mRNA for MLH1 protein, partial.
ACCESSION AJ012747
VERSION AJ012747.1 GI:3893080
KEYWORDS MLH1 gene; MLH1 protein.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 2381)
JOURNAL Mol. Gen. Genet. 262 (4-5), 633-642 (1999)
MEDLINE 20092468
PUBMED 10628846
REFERENCE 2 (bases 1 to 2381)
AUTHORS Jean, M., Pelletier, J., Hilbert, M., Belzile, F. and Kunze, R.
TITLE Isolation and characterization of AtMLH1, a Mult. Homologue from
Arabidopsis thaliana
JOURNAL Submitted (17-NOV-1998) Kunze R., Institute of Genetics, University
of Munich, Maria Ward-Strasse 1a, 80638 Muenchen, GERMANY
FEATURES
Location/Qualifiers

```

```

source 1..2381
/organism="Arabidopsis thaliana"
/variant="Columbia"
/db_xref="taxon:3702"
3..2216
/gene="MLH1"
<3..2216
CS
/gene="MLH1"
/function="DNA mismatch repair"
/codon_start=1
/product="MLH1 protein"
/protein_id="CAI10163.1"
/db_xref="GI:3893081"
/translation="MIDSSILPAEMEEESPATVIREPPKIORLEESVYVNR
VIOBPVSAVEIENSLADSSISVYVNDGKILQVSDPDGIVREDDPILCEHT
TSKILNREDFLSMGRFREALASMTYVAVYVITTTGQIHGYVSVSDCMHEP
KACAAVGTQIVENLEFNMIAKRTLONSADYIVLIDLSMAIHYNVFSCKH
GAVKADVSVSPSRIDSIRSYGVSAKMLKASPEVYMSINPREHVDINHT
AKTIVILFINDLVECSALKRAIEIYVATLTPKASPEVYMSINPREHVDINHT
KREYSLNOEITTEMIOSEVEYELRNANDRFOBOKEVYIOSTLSOKSDSVOKP
SGKOTKVPYVKKVVRIDSDPAGRLHAFLOPKOSLPDKVSSLYVRSVROBRNKE
TADLSYOEILAIYDSCHEGMEYVANCYVGMADDPALVYVNTHIYANVINSK
ELMYQTLRFPAHFNALQSDPAPLSLILALKEEDLPDNDKDLERLAEATE
LKEKAEMLREYVSHIDSSANLSRLPVLIDYTPMDRPEFLDLGNDVEWDEKS
CFQVSAIAGNEFYAMHPPLIPNSGDIGQYSRGSOSORSDENVDMDLDDL
LSDENMAQREMSIOHVLEPPSRFLFKPPASMASNQTGVKSLKIKYIFERC"
BASE COUNT 721 a 450 c 548 g 662 t
ORIGIN
Query Match 37.0%; Score 926.2; DB 8; Length 2381;
Best Local Similarity 65.4%; Pred. No. 2.2e-178;
Matches 1408; Conservative 0; Mismatches 733; Indels 12; Gaps 3;
158 GGGAGCCCCCGCATCCGAGGTTGGAGAGTGGTGGTGAACCGCATCGCGGGGGG 217
73 GAGACCAACCGAAGATTAACGCTTAGAAGATCAGTACACCGTATCGCAGCTGTG 132
218 AGGTGATCCAGCGCGCTGCTGCGGCGTGAAGAGACCTCATCGAAGACGCTGACG 277
133 AAGTATCAAGGTCACATTCCTGAGCTGGAAGAGCTGTTGAGAAAGCTCGACGCG 192
278 GCGCTCCAGGCTCCGCTTGGCGTGAAGAGCGTGGCTCAAGCTCATCCAGCTCCG 337
193 ATTCAAGTTCATTAAGCGCTGTTGCAAGAGCGGTTGAAACTCATTAAGCTCCG 252
338 ATGACGCCATGCGATCAGGTTTGAGAGTATGCAATATGTCGCAAGAGCATTAAG 397
253 ACGAGCGTCACGCTATTAGAGCTGAAAGACTTCCGATGACTATGCGAGACATACA 312
398 CAAAGTATCTGCATACAGAGATCTGCAAGACCAVAAAATCGATGGGCTTCAGAG 457
313 CGAAGCTGACATTAAGTGAAGATTTGCTCTGCTGATCAATGAGATTTAGAGAG 372
458 CTTGGCTAGTATGACTTATGTTGGCATGTACCGTGCACAGATCAAGAGAGCCAT 517
373 CATTTAGTATGATGACCTATGTTGCTCATGTTACAGTGAATCTATTAAGGCCA 432
518 TGCACGGCTACAGGCTTTCTACAGAGATGTTGATGAGATGAGATGAGCCCTGCG 577
433 TTCAATGTTATGATGATGCTTTATAGAGATGCTGATGAGATGAGCAACAAAGG 492
578 CTGCGGTGAAGCAACTCAAGTATGTTGAAATCTATTTTCAACATGATGAGCCCA 637
493 CTGCTGTCAAGAGAACACAGATATGCTGAGAAATTTGTTTCAATATGATGCTA 552
638 AAAAAATTCAGAACTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 697
553 GGAAGACACTTAAATTTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 612
698 TTGACATCATCAATCAAGTATTAATCTCTTTCAGAGAAAGATGAGCAATAGAGAG 757
613 TGGCTTTCATTAACAATATGTCAGCTTTCTTGTGCAAGAGATGAGAGCTGTAAGCG 672

```

QY 758 ATGTTATAGTCGACAGTACATCTCAAGGTTAGTCTATCAGAGAGTCTATGAGGCTT 817
 DB 673 ATGTTACTAGTCTGCTACCTTCAAGGCTTGATCAATAGTGTATATGAGTAT 732
 QY 818 CTGTCCTGCTGATCTATAGAAATTAAGGTTTATGAGAGTGTGAGATTTCAATCT 877
 DB 733 CAGTTGCAAGAACTTGAAGTATAGAGTCTCTCTGACTCTCTGCTGCTT 792
 QY 878 TCAAGATGATGTTTACATCTCAATATATGTCGAAAGAGATTAATCAATGATTC 937
 DB 793 TTGATATAGAGGTTTATATTCATATTCATATGTTGCTAGAGAACTATATGCTGC 852
 QY 938 TTTTCAATTAATGATAGGCTTGTAGACTGTACTCTGTTGAAAAGAGCTATTTGATTTGCT 997
 DB 853 TTTTCAATTAATGATAGTGTGTGAGATGCTCTGCTTAAAAGAGCATTAATTTGTTT 912
 QY 998 ACTCTGCAATGTCCTCAAGATCCCAACCTTTCATATATCATCTCATATCTTCCAT 1057
 DB 913 ATGCTGCAACATTTGCCAAAAGCATCAAAACCTTTTGTCTCATGTCATCAATTTTGCAC 972
 QY 1058 CAGAACATGATGTTTATATATACCCCAACCAAGAAAGAGTTAGGCTTTGATCAAG 1117
 DB 973 GGGAAACATGTTGATATATATTCACCCCAACCAAGAAAGAGTTAGGCTTTTAAACGAG 1032
 QY 1118 AGCTATTTATTAACAATTAAGAAATGCTATTTGAGAAAGAGCTGATTAATTAACA 1177
 DB 1033 AAATCATTTATTTGATGATGATACAGTCAAGAGTTGAAATTAACAGAAACGCAATGATA 1092
 QY 1178 CAGAGATATTCCAACACGAGCATTAACATATACAGGATTTGCTCAACGTCACACAAA 1237
 DB 1093 CTAGGACGTTTCAAGACCAAGAGTGAATTAATCAATCTACGTTTAACTCTAGAAA 1152
 QY 1238 AGATTAAGTTTCTGAGGCGAGTATGGTTCTGAGAACAAAATCTCAAAAATTTCTGTGA 1297
 DB 1153 GTGAT---TCTCCAGTTTCTCAGAGAGCTTCTGACAAAAGACACAAAGTTCTCTGTA 1209
 QY 1298 GCCAATGCTGACAGATCCACGCAATCTGGAAGATTTGACACCTACTGCGACG 1357
 DB 1210 ACAAAATGCTGAGAACAGATTCATCAATCAGCTGGAAGGTTACATGCTTTTGGAAC 1269
 QY 1358 GGCATCTTCAATCT---TGAAAAGAAATTTGATCTTGTATCTGTAAAGAAATTTGTAA 1414
 DB 1270 CCAAGCCACAAAGTCTCTCCACAAAGTTCTTCAATTTGAGTATGATAGGCTCTCTTAA 1339
 QY 1415 GATCAGAGAGAAACCAAAAAGATGCTGATTTGTCAAGCCGTCATGAGCTCTGTGG 1474
 DB 1330 GGCAGAGAGAAACCAAAAAGATGCTGATTTGTCTAGTGTCCAGAACTTATTTGCTG 1389
 QY 1475 AAATGATTTAGCTTCCATCTGCTGCTTTTGGACATTTGCAAGACTGCACATATGTTG 1534
 DB 1390 GAGTTGACAGCTGCTGACATCCAGGTATGCTGAGACTGTAAAGAAATTTGACATATGTTG 1449
 QY 1535 GACTTGCCGATGAAAGCTTGTGATTAACAACATTAACCCGTTTCTTACTCTTGAATG 1594
 DB 1450 GAATGCGAGATGATTTTGTCTTACTTCAATTAACCCATCTATATCTTACCAATG 1509
 QY 1595 TGTAAATATTTAGTAAGAACTTATGTACAGCAAGCTTTGCGCTTTTGGAACTTCA 1654
 DB 1510 TGTGATATCTCAGCAAGAGCTATGATACAGCAAACTCTTCTGCTTTTCTCTATTTTA 1559
 QY 1655 ATGCTATTAAGCTCAGTGAACACAGCTCCACTTCAGAGATTTGCTGATGAGCACTGAAG 1714
 DB 1570 ACCGAAATACAGCTTACGATCCAGCCCTTTGTGAGAGTTGATTTGCTGCTGAAAG 1629
 QY 1715 ACGATGATTT-----GATGATGATGAAAAAGATGATGAGAAATGAGATTTGCAAG 1768
 DB 1630 AGGAGGATCTAGATCCAGAAATATATCAAAAAGATGATGAAAGAAAGATTTCTGAAA 1689
 QY 1769 TAAACACTGAGATTAATAAGAAATATCTGATGATTAATGATTAATCTTTCTATTACA 1828
 DB 1690 TGAATACAGAACTCTCTCAGAGAAAGAAATGTTTGAAGAGATTAATTTTCAAGCTGCA 1749

QY 1829 TTGATCAAGATGCGAAATGACAAAGACTTCTGTTGACTGAGACCACTGACCCCTGATA 1888
 DB 1750 TTGACTCCAGTGCAGAAATTTGTCAAGGCTTCTGTGATATCTGACAGATATACCTGACA 1809
 QY 1889 TGGACGCTTCTCCAGAAATTTGTGTGCTTTAGAGAAATGATGATTTGATGATGACAGA 1948
 DB 1810 TGGATGCTGTTCCGAAATTTTACTATGCTTGGGAAATGATGATGAGGAGAAATGAGA 1869
 QY 1949 AAGATGCTTCCAGAAAGATGACTTCTGCTGAGAGAAATTTCTATGACCTTATCCCCAA 2008
 DB 1870 AAGATGCTTCCAGAAAGATGACTTCTGCTGAGAGAAATTTCTATGACCTTATCCCCAA 1929
 QY 2009 TCTTCCAAATTCATCTGCGAATGATGATTTATACAAAGAAATGATGATTAATG 2068
 DB 1930 TTTTCCAAACCCATCCGCTGACGCTGATTCATGTTATGAGAGAGTGAAGCTCTC 1989
 QY 2069 CTGATGAACATGCTGAGATATCTAATATACATGAAATGAGCTGATGATTAACACTC 2128
 DB 1990 AGGAAAGTCAATTTAGAGGTTAAGCTGATATGAGAGACAACTTGACCAAGATCTC 2049
 QY 2129 TTGCGAAGCAGAAAGCAGATGAGGCCCCCAACGTCAGTGTGACATGCTTGTTC 2188
 DB 2050 TGTGAGATGCTGAAAGAGCAGATGAGGACAGCTGATGATGATCAATCCACAGCTGTTTC 2109
 QY 2189 CATCATGCGACTTCTCTCAAGCCCGGAGATGATGAGCAAGATGAGAACTTTGTC 2248
 DB 2110 CGTCAATGAGATTTGCTTGAAGCCACAGCTTCCATGCTTCAATGAGGACTTTGTA 2169
 QY 2249 AGTGTGCTCTGAGAAACTCTACAAAGATTTTGAAGGTTTACTGCTA 2301
 DB 2170 AGTGAAGATCCCTTGAAGAAAGCTGTACAAAGATTTGCAAGATGCTTAACGTAA 2222

RESULT 5
 AP003238
 LOCUS 160704 bp DNA linear PLN 13-APR-2002
 DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
 PAC clone: P0401G10.
 ACCESSION AP003238
 VERSION AP003238.4
 KEYWORDS GI:20146211
 SOURCE
 ORGANISM
 Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,
 clone: P0401G10.
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
 PAC clone: P0401G10
 Published only in Database (2001)
 2 (bases 1 to 160704)
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Direct Submission
 Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
 Agrobiological Sciences, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)

COMMENT
 On Apr 12, 2002 this sequence version replaced gi:17933034.
 Genes were predicted from the integrated results of the following:
 GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
 (October 1998 version). The genomic sequence was searched against
 NCBI Nonredundant Protein database, nr
 (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
 RCP. Protein homologies of the coding regions were searched against
 NCBI Nonredundant Protein database with BLASTP2.0. ESTs represent
 the identified cDNA sequences using BLASTN 2.0 with the
 corresponding DBJ accession no. and RCP clone ID.
 A gene with identity or significant homology to a protein is
 classified based on the protein name to indicate the homology level
 such as same name, 'putative-' and '-like protein'. A gene without

significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein. The orientation of the sequence is from T7 to SP6 of the PAC clone. This sequence of P0401G10 clone has an overlap with O11294.F06 clone (DBJ: AP004326) at 5' end and an overlap with P0483G10 clone (DBJ: AP003263) at the position 112,620 to 160,704 of 3' end. The sequence of this clone ends at the position 48,085 of P0483G10. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://ftp.dna.affrc.go.jp/Genomeseq.html>.

FEATURES

source
1. 160704
/organism="Oryza sativa (japonica cultivar group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="1"
/clone="P0401G10"
join(3181..4553,4587..4623)
/gene="P0401G10.1"
join(3181..4553,4587..4623)
/gene="P0401G10.1"
/note="contains ESTs"
C26818(C50145), C22453(C62707), C22454(C50145),
C96934(C62707)
/codon_start=1
/product="protein kinase-like"
/protein_id="BAB88994.1"
/db_xref="GI:20146212"
/translation="MRHALLGVFALCLLRQAOPDYPTAKPSTTWANTDALHH
VATYGSVARALLRLPARLGGSPFAFGFECNHGACAFELGLVAVVYCNAGSI
TAVTGTIPQVANSANRAPVGCATAEITADSDIYLRPGGVLMASAGACGVGMS
INDSGLVLEFDSNRTWQSFDPIDTILVGGSLQCARLTANASPSDSRRTIYLA
ADGGLAAYDAKPPQRYVYLYGSNAGAAATNGSLAYLDRPGGQALATQVLPVAA
GYQVRLRHDHRLLYEWSNMGMEATGDLHPYGGCAVPTVCGALVGTCTMCS
CPDAANFAYDFRRPGCGVPTSPAPATCSRRARHRLVSLRDTAFNSDTMRT
LRYGAAKACACACGACMAAFVYGFDPNGFCYLGSEVSLTETMPEVYHNSMS
HRIAGSILTYTKEKSL"
complement(join(5066..5279,5401..5468))
/gene="P0401G10.2"
complement(join(5066..5279,5401..5468))
/gene="P0401G10.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB88995.1"
/db_xref="GI:20146213"
/translation="MNFSLHYGWTYVAKGAREELWQIFNPSGNOESAGRGIGS
AAEERRRRRGRGSRSGDARGTOTRRRRRVPSPDMMTGRRPKS"
join(6174..6277,6665..6986,7334..7735,7878..8222,
8572..8676)
/gene="P0401G10.3"
join(6174..6277,6665..6986,7334..7735,7878..8222,
8572..8676)
/note="hypothetical protein similar to Arabidopsis thaliana chromosome2, At2g28480"
/codon_start=1
/protein_id="BAB88996.1"
/db_xref="GI:20146214"
/translation="MATRVLSQNRKIASFLINLISORAPISPEPLRPAISFGC
LNPPYQSPPMVAMASHSNLYLSDDGPKREIEVEVSKRGRTILKRLQRR
EKRRKANKNDPRIRKGRKIKQKPTPEAKRTYTERAKLEAMLVAKKYEIA
KAQPGMAKODLDEERFYLKVQSKNSNYPVGRGVFGVILNMLHMKHETVAV
ICKCKPGIOEIVASEIARLGGIPINTIGNDIYFGRKNVQVOPDMVLSKTK
ALEKSYEOSLETYRRRIAYSEKLELYHIALYGNOSONDPYVCDRRASLSKM
LEPSQGLSHMDNGSDITDASEDEANPSYDNDDETGNICTVLYDGGFRL
IVHAKTSVGRYSQDLKLVVAGEOYHEKLELSA
join(12510..12707,13071..13212,14284..14395,16176..16212,
16496..16849)
/gene="P0401G10.4"
join(12510..12707,13071..13212,14284..14395,16176..16212,
16496..16849)
/note="DNA mismatch repair protein"
/codon_start=1

misc_feature

gene
/gene="P0401G10.4"
/note="contains EST C27216(C51353)
similar to RNA-binding protein"
/codon_start=1
/protein_id="BAB88997.1"
/db_xref="GI:20146215"
/translation="MAVSAGGGGAGAGFGDTTLKRVFVGLAMETOREGMRGPEQ
GILELVITDNTKNSKSGYGFVTFREPAARACFDEPVYIDGRANCLAYLQVQ
SKRAASLQPYAGHRRAMKSIQOTGGGASISMAHGLQOGLPTPNVYGYSPYSDYG
YPSYQANGYCGAGQOMRACAGAGAGITMANNTGGGIVSGYSPYQYPAVAAA
GTAQWYPMQYTAALAAAGVAPTATTAISQTLTANVAGLQVAGVTAALPNSYGOAGL
SEMF"
complement(19970..23375)
/gene="P0401G10.5"
complement(19970..23375)
/gene="P0401G10.5"
/note="probably inactive due to frame shift(s) in CDS
pseudogene, similar to GDSL-motif lipase/acylhydrolase"
/pseudo
join(25336..25599,25803..26041,26224..26428,26676..26744,
26826..27012,27717..27883,28412..29002)
/gene="P0401G10.6"
join(25336..25599,25803..26041,26224..26428,26676..26744,
26826..27012,27717..27883,28412..29002)
/gene="P0401G10.6"
/note="contains ESTs
C73071(E2861), D46758(S11633), A0032617(S11633)
similar to Arabidopsis thaliana T22A6.120
unknown protein"
/codon_start=1
/protein_id="BAB88998.1"
/db_xref="GI:20146216"
/translation="MAKRAKQVSESARSICGLVDIANDIRLAKCKQSGPDLIL
ELDADKODIVLPGNLTIVGVSKSIKDKCKERRRPSDVLTFQOMSEKRELSLQK
IPSGFNAFEKTCQKQDASIKRSLAFDMCTLTLYVA,LSKHAILIKQHVAVST
MEPALARFKETGTHIVYCKMGKQVILYIKQNSLQADVAVQRLKEMSPQRLD
ANGSDLSIADSVAKDKVARREORLFEVSNLQNSYSSNEELVAMPKRGSGDKDII
SHSWLTVQAEPRVYINSTPIRSLINYPGCGFNLHNLILFKRPREELHOLE
FQLRWANVAYSDILPQRRKQSSASLPLNLIQPKLYCTNMIIQLEDYTPVQVPE
AEIKRYEPEISNMRFSHVCTAPVDSQSSITVGAAHLEVSRPKLILRLHFSVC
NATSVKPEWMSGPNLQCKSLSTLSTHSTPAALPAPRPAEVLNIVYGGGPV
PVQPKLIRFVDTEMLRGDPLDGYWVVGAKLHRRKISLRVXISLITVNSPDE
FSPDEER"
complement(join(29649..29810,30106..30253,30610..30692,
31728..31819,31966..32037,32133..32223,32519..32559,
32678..32899,32926..32974))
/gene="P0401G10.7"
complement(join(29649..29810,30106..30253,30610..30692,
31728..31819,31966..32037,32133..32223,32519..32559,
32678..32899,32926..32974))
/gene="P0401G10.7"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB88999.1"
/db_xref="GI:20146217"
/translation="MEGLVLPNLFENKRVGGFYHNLPAISCPVLVPSNGMCLRAP
DHRSPFTIRAAALAAAPLIPGNRGSGDERDAIIPNCSRSRISMTGSYKCAVDFK
AQAATIGSPADPFLIRAKPEVHVLVPGNPGIYAFKDFVEELYNLGGQAVITVA
ADPHKTHNRMSYDPCFRCKMNLITRYGVNLGGLVPELILNKKSMKQSAIGTIAAS
SLSKGVSTFSYFISLQASVTRGIMRLILPSSSVYAVETGHLMLSEPPMNF
SNKQDQIFLFEVDHMGPLAHLEEVNSFKATKILLIYRS"
join(33330..33566,33690..33862,33943..34015,34105..34239,
34309..34519,34843..34991,35126..35224,35874..35946,
36133..36265,36557..36651,36731..36839,37093..37252,
37390..37551,37771..37942,38211..38377)
/gene="P0401G10.8"
join(33330..33566,33690..33862,33943..34015,34105..34239,
34309..34519,34843..34991,35126..35224,35874..35946,
36133..36265,36557..36651,36731..36839,37093..37252,
37390..37551,37771..37942,38211..38377)
/gene="P0401G10.8"
/note="DNA mismatch repair protein"
/codon_start=1

CDS

/gene="hmlh1"
42..2312
/function="DNA mismatch repair"
/note="human homolog of E. coli mult. gene product,
Swiss-Prot Accession Number P23367"
/codon_start=1
/protein_id="AA17374.1"
/db_xref="GI:466462"

BASE COUNT 723 a 539 c 599 g 642 t
ORIGIN

Query Match 12.9%; Score 321.4; DB 9; Length 2503;
Best Local Similarity 57.8%; Pred. No. 4.7e-55;
Matches 596; Conservative 0; Mismatches 426; Indels 9; Gaps 1;

OY 172 ATCCGAGGTTGGAGGAGCGGTGGTGAACCGGATCGGGGGGAGGTGATCCAGCGG 231
DB 63 ATTCGGGGGCTGGAGGAGAGAGTGGTGAACCGGATCGGGGGGAGGTGATCCAGCGG 122
OY 232 CCGTCTCGGGGTGAAGAGCTCATCGAGACAGCTCGACCTGGCGCTCCAGCGTC 291
DB 123 CCAGCTATGTTATCAAGAAGATGATGAGACCTGTTAATGCAAAATCCAAAGTAT 182
OY 292 TCCGTGGGTTGAAGAGCGGTGGTGAACCGGATCGGGGGGAGGTGATCCAGCGG 351
DB 183 CAGGTGATGTTGAAGAGGAGGCGCTGAAGTGTGATGATCCAAAGCAATGGCACCGG 242
OY 352 ATCAGGTTGAGATTTGGCAATATTCGCAAGGACATACCTACCTCAAGTTATTCGCA 411
DB 243 ATCAGGAAGAAGATCGATGATTTGTATGTAAGGTTCACTACTAGTAACATCGACGTC 302
OY 412 TACGAGATCTGCAGACCATATAATCGATGGGTTTCAGAGGGGAGGCGTTGGCTAGTATG 471
DB 303 TTTGAGATTAGCCAGATATTTTACCTATGCTTTCGAGGTGAGGCTTTGGCCAGCAT 362
OY 472 ACTTATGTGGCATGTTACCGTGAAGATTAACAGAGGCAATTCGACGGCTACAGG 531
DB 363 AGCCATGAGGCTCATGTTTCATTCACAGGAAGAGTGAAGAGTGTGATGATACG 422
OY 532 GTTCTTACAGAGTGTGTAATGAGAAATGAGCTTAAGCCTTGGCGGGGAGTAAAGGA 591
DB 423 GCAATTTACTCGATGAGAAATGTAAGCCCTCCTTAACATGAGCTGGCAATCAAGG 482
OY 592 ACTCAAGTATGTTGAAGAAATCTATTTTACAACATGTAAGCCGCAAGAAACATTCAG 651
DB 483 ACCCAATACGCTGAGAGACCTTTTACAACATGACGAGAGAGAAAGCTTTAA 542
OY 652 AACCTCATGATGACTACCCCAAGATCGTACCTTCATCAGTCGTTGGCTGCATCAG 711
DB 543 AATCCAAAGGAAGATATGGAAGAAATTTTGAAGTGTGGAGATATTCAGTACAT 602
OY 712 ATCAAGTTACCTCTCTGCAAGAAAGCATGGAGCAATAGAGCAAGATTTTATGCA 771
DB 603 GAGGATATGTTCTCAGTTAAACAAAGAGAGAGAGAGTGTGATGAGACACATA 662
OY 772 AGTACATCTCAAGTATGATCAGAGAGTGTCTATGAGGCTTCTGCTTGTGAT 831
DB 663 CCCAATGCCCAACCGGAGCATATTCGCTCCGCTTGGAAATGCTGTAGTGGAGAA 722

OY 832 CTCATGAAATTAAGTTTCATATGAGATGCTGCAGATTCATCTTCAGATGATGCT 891
DB 723 CTGATGAAATGATGATG-----GAGATTAACCCCTACGCTTCAAAATGATGCT 773
OY 892 TACATCTCAAGTCAAAATATGTTGGCAAGAGATTTACATGATTTCTTTATTAATGAT 951
DB 774 TACATTTCAAGCAAACTACAGAGAGAGTGTGATGATGATGATGATGATGATGAT 833
OY 952 AGCTTTGAGTGTACTGCTCTTGAAGAGAGATTTGATGATGATGATGATGATGATG 1011
DB 834 CGTCTGTGAAATCAACCTTCCTGAGAAAGAGATGAAAGAGATGATGATGATGATGAT 893
OY 1012 CCGCAAGATTCACCACTTTCATATGATGATGATGATGATGATGATGATGATGAT 1071
DB 894 CCGCAAGATTCACCACTTTCATATGATGATGATGATGATGATGATGATGATGAT 953
OY 1072 GTTAATATACCCCAACCAAGAGAGGTTAGCTTTGATATCAAGAGGCTTTATGAA 1131
DB 954 GTTAATGTGACACCCCAAGAGATGAAAGTCACTTCTCAGAGAGAGAGATGATGAT 1013
OY 1132 ACAATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1191
DB 1014 CGGTGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1073
OY 1192 ACTCAGGATTT 1202
DB 1074 ACCCAGACTTT 1084
RESULT 7
AR122159 2484 bp DNA linear PAT 16-MAY-2001
LOCUS AR122159
DEFINITION Sequence 4 from patent US 6165713.
ACCESSION AR122159
VERSION AR122159.1 GI:14106476
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2484)
AUTHORS Liskay,R.M., Bronner,C.Eric., Baker,S.M., Bollag,R.J. and
TITLE Composition and methods relating to DNA mismatch repair genes
JOURNAL Patent: US 6165713-A 4 26-DEC-2000;
FEATURES
source 1..2484
/organism="unknown"
BASE COUNT 720 a 535 c 594 g 635 t
ORIGIN
Query Match 12.8%; Score 319.8; DB 6; Length 2484;
Best Local Similarity 57.7%; Pred. No. 1e-54;
Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;

OY 172 ATCCGAGGTTGGAGGAGCGGTGGTGAACCGGATCGGGGGGAGGTGATCCAGCGG 231
DB 43 ATTCGGGGGCTGGAGGAGAGAGTGGTGAACCGGATCGGGGGGAGGTGATCCAGCGG 102
OY 332 CCGTCTCGGGGTGAAGAGGCTCATCGAGACAGCTTCGAGCGGCGCTTCAGCGCTC 291
DB 103 CAGCTAATGCTATCAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 162
OY 292 TCCGTGGGTTGAAGAGCGGTGGTGAACCGGATCGGCTTCATCAGTCTCCGATGAGGCGCATGGC 351
DB 163 CAGTGTATGTTAAGAGGAGGCGCTGAAGTGTGATGATGATGATGATGATGATGAT 222
OY 352 ATCAAGTTGAGATTTGGCAATATTCGCAAGGCAATAGAGCAAGTATATGCA 411
DB 223 ATCAGGAAGAAGATTCGATGATTTGTATGTAAGAGTGTACATCTAAGTATGCAAGCTC 282
OY 412 TACGAGATCTGCAGACCATATAATCGATGGGTTTCAGAGGGGAGGCTTTGGCTAGTATG 471

Db 283 TTGAGGATTGACGAGTATTTCTACCTATGCTTCGAGTGAGGCTTTGGCCACATA 342
 Qy 472 ACTTATGTTGGCCATGTTACCGTGACACAGATTAACAGAAAGCCAAATGACGGCTACAG 531
 Db 343 AGCCATGAGGCTGATCTATCTATTAACAGCAAAACAGCTGATGAGAGTGATGATCAGA 402
 Qy 532 GTTCTTACAGAGATGATGATGAGAAATGAGACCTTACGCTTGGCTGCGGTGAAGGA 591
 Db 403 GCAAGTACTACGATGAGAAACCTGAAAGCCCTTAAACATGCTGCTGCAATCAAGG 462
 Qy 592 ACTCAAGTATGAGTAAATGATTTTCAACATGAGTGGCCGCAAGAAACATGAGG 651
 Db 463 ACCGATATACGATGAGAGACCTTTTTCACATGAGCCAGAGAGAAAGTTTAA 522
 Qy 652 AACCTCAATGATGATGAGCCCAAGATGATGAGCTTATGATGAGTGGCTGAGTCAATC 711
 Db 523 AATCCAGTATGAGTAAATGAGAAATTTTGAAGTTGTTGAGGATGATGATGATGAG 582
 Qy 712 ATCAAGTATGATGATGAGAAAGATGAGACCAATGAGAGATGATGATGATGAG 771
 Db 583 GAGGATGATGATGATGAGTAAAGAAACAGAGACAGATGATGATGATGAG 642
 Qy 772 AGTACATGCTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 831
 Db 643 CCCATGCTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 702
 Qy 832 CTCATGAGAAATGAGTATGATGATGATGATGATGATGATGATGATGATGAT 891
 Db 703 CTGATGAGAAATGAGTATGATGATGATGATGATGATGATGATGATGATGAT 753
 Qy 892 TACATGCTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 951
 Db 754 TACATGCTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 813
 Qy 952 AGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1011
 Db 814 CGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 873
 Qy 1012 CCTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1071
 Db 874 CCCAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 933
 Qy 1072 GTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1131
 Db 934 GTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 993
 Qy 1132 ACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1191
 Db 994 CGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1053
 Qy 1192 ACTGAGCAT 1202
 Db 1054 ACCGACATTT 1064

RESULT 8
 LOCUS AR131133 2484 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 4 from patent US 6191268.
 ACCESSION AR131133
 VERSION AR131133.1 GI:14119458
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2484)
 AUTHORS Listay R.M., Bronner, C.Eric., Baker, S.M., Bollag, R.J. and
 TITLE Koldner, R.D.
 JOURNAL Compositions and methods relating to DNA mismatch repair genes
 FEATURES Patent: US 6191268-A 4 20-FEB-2001;
 location/Qualifiers
 1..2484
 source /organism="unknown"

BASE COUNT 720 a 535 c 594 g 635 t
 ORIGIN
 Query Match 12.8%; Score 319.8; DB 6; Length 2484;
 Best Local Similarity 57.7%; Pred No. 1e-54;
 Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;
 Qy 172 ATCCGAGGTTGAGAGAGTGTGTGTAACCCGATGCGGGGGAGTATGATCAGCG 231
 Db 43 ATTGCGGGGCTGAGAGAGAGTGTGTGTAACCCGATGCGGGGGAGTATGATCAGCG 102
 Qy 232 CGCTGCTGGCGGTGAGAGAGTGTGTGTAACCCGATGCGGGGGAGTATGATCAGCG 291
 Db 103 CCAGTATGCTATCAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 162
 Qy 292 TCCGTTGGGTTGAGAGAGTGTGTGTAACCCGATGCGGGGGAGTATGATCAGCG 351
 Db 163 CAGTATGCTATCAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 222
 Qy 352 ATCCGAGGTTGAGAGAGTGTGTGTAACCCGATGCGGGGGAGTATGATCAGCG 411
 Db 223 ATCCGAGGTTGAGAGAGTGTGTGTAACCCGATGCGGGGGAGTATGATCAGCG 282
 Qy 412 TACGAGATGCTGACAGCATTAAATGATGAGGCTTCAAGGGAGGCTTGGCTAGTATG 471
 Db 283 TTTGAGGATTTTACCGATGATGATGATGATGATGATGATGATGATGATGAT 342
 Qy 472 ACTTATGTTGGCCATGTTACCGTGACACAGATTAACAGAAAGCCAAATGACGGCTACAG 531
 Db 343 AGCCATGAGGCTGATCTATCTATTAACAGCAAAACAGCTGATGAGAGTGATGATCAGA 402
 Qy 532 GTTCTTACAGAGATGATGATGAGAAATGAGACCTTACGCTTGGCTGCGGTGAAGGA 591
 Db 403 GCAAGTACTACGATGAGAAACCTGAAAGCCCTTAAACATGCTGCTGCAATCAAGG 462
 Qy 592 ACTCAAGTATGATGATGAGCCCAAGATGATGAGCTTATGATGAGTGGCTGAGTCAATC 711
 Db 463 ACCGATATACGATGAGAGACCTTTTTCACATGAGCCAGAGAGAAAGTTTAA 522
 Qy 712 AGTACATGCTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 831
 Db 523 AATCCAGTATGATGATGAGAAATGAGAAACCTGAAAGCCCTTAAACATGCTGCTGCAAT 582
 Qy 832 CTCATGAGAAATGAGTATGATGATGATGATGATGATGATGATGATGATGAT 891
 Db 703 CTGATGAGAAATGAGTATGATGATGATGATGATGATGATGATGATGATGAT 753
 Qy 892 TACATGCTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 951
 Db 754 TACATGCTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 813
 Qy 952 AGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1011
 Db 814 CGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 873
 Qy 1012 CCTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1071
 Db 874 CCCAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 933
 Qy 1072 GTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1131
 Db 934 GTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 993
 Qy 1132 ACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1191

Db 994 CGGGTGCAGCAGCATGAGAGCAAGCTCCTGGGCTCCATTCCTCCAGATGACTTC 1053
OY 1192 ACTAGGCATT 1202
Db 1054 ACCCAGACTTT 1064

RESULT 9
AX214172
LOCUS AX214172 2484 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 10 from Patent WO0159092.
ACCESSION AX214172
VERSION AX214172.1 GI:15524260
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Nicolaiides, N.C., Sass, P.M., Grasso, L., Vogelstein, B. and
Kinzler, K.W.
TITLE Methods for generating hypermutable microbes
JOURNAL Patent: WO 0159092-A 10 16-AUG-2001;
The Johns Hopkins University (US)
FEATURES
source 1. 2484
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 720 a 535 c 594 g 635 t
ORIGIN

Query Match 12.8%; Score 319.8; DB 6; Length 2484;
Best Local Similarity 57.7%; Pred. No. 1e-54;
Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;

OY 172 ATCCGAGGTTGGAGAGTGGTGTGAACCCGATCGGGCGGGGAGGTATCCAGCGG 231
Db 43 ATTCGGCGCTGGAGAGACAGTGTGAACCCGATCGGGCGGGGAGGTATCCAGCGG 102
OY 232 CCGTGTGGCGGTGAAGAGAGTGTATCCAGAACACCTTCGAGCGGCGCCGCCGCGC 291
Db 103 CAGCTAATGCTATCAAGAGATGATGAGAACTGTTAGTCAAAATTCACAAATATT 162
OY 292 TCCGTTGGGTTGAAGAGAGTGGCTCAAGCTCATCGAGTCTCCGATGACGCCATGCG 351
Db 163 CAAAGTATGTTAAAGAGAGAGCGCTGAAGTGTATGATCCAGAACATGCGACCGG 222
OY 352 ATCAGGTTGAGGATTTGGCAATATTGGCGAAGAGCACTACTCAAGTTATCTGCA 411
Db 223 ATCAGAGAAAGAGATCTGATGTATGTATGTAAGAGTTCACTACTAGTAACTGCAATCC 282
OY 412 TACGAGGATTCGCAACCATTAATCGATGGGGTCAAGGGGAGGCTTTGGCTGATATG 471
Db 283 TTTGAGGATTTAGCCAGTATTTCTACCTATGAGCTTTCGAGGAGGCTTTGGCCACGATA 342
OY 472 ACTATGTTGGCCATGTTACCTGACAAAGATACAGAAAGGCCAATTGACGCGTACAGG 531
Db 343 AGCCATGTGGCTCATGTTACTTATACAAAGAAAGCAAGCTGATGAGAAAGTGTGATACAGA 402
OY 532 GTTCTTACAGAGATGTGTATATGAGAAAGAGCTTAAGCCTTGGCGTGGTGAAGA 591
Db 403 GCAAGTACTACATGAGAAAGTGAAGCCCTCTTAAGCAATGCTGTGCAATCAAGAG 462
OY 592 ACTCAAGTCATGTTGAAGAAATATATTTCACAACTGTCAGCCCGCAAGAAAGATTTGCG 651
Db 463 ACCCAGATCAGGTGAGAGACCTTTTTCACAACTGTCAGCCCGCAAGAAAGATTTTAA 522
OY 652 AACCTCAATGATGATACCCCAAGATCGTACACTCATGAGTGGCTTCCAGTCCATCAG 711
Db 523 AATCCAGTGAAGATATGAGAAATTTTGGAGTGTGGCAGAGTATTCAGTACAGAAAT 582
OY 712 ATCAACGTTACCTCTCTTGGAGAAAGATGAGCCCAATAGAGAGATGATGATGATGA 771

Db 583 GCAGGCATTAATTTCTAGTAAAAAACAGAGAGACAGTAGTGAATGTTAGACACTA 642
OY 772 AGTACATCTCAAGATTAGATGATGAGAGTGTCTATGAGGCGCTTCTGCTGCGAT 831
Db 643 CCCAATGCCCTCAACCGTGAAGCAATATTCCTTCATCTTTGAAATGCTGTAGTGAAGA 702
OY 832 CTCATAGAAATTAAGCTTTCATATGAGGATGCTGAGATTCATATCTTCANATGATGATG 891
Db 703 CTGATAGAAATTTGATGT-----GAGATTAAGCCCTTCAAGTCAATGATGAT 753
OY 892 TACATCTCAATTAAGCAATTAATGTCGCAAGAAAGATATTCATATGATCTTTTATTAATGAT 951
Db 754 TACATATCAATGCAAACTACTGATGAGAAAGAGCACTCTTACTCTTCATCAACCAT 813
OY 952 AGGCTGTAGACTGATGCTGCTTTGAAAGAGCTATGATTAATTTGTACTCTGCAACATTTG 1011
Db 814 CGTCTGTAGATCACTCTCTTGGAAAGCCATAGAAAGATGATGACGCTATTTG 873
OY 1012 CCTCAAGCATCCAAACCTTTTATATATGATGCTATGATCTTCATCAGAAACAGGTGAT 1071
Db 874 CCAAAACACACACACCATCTCTGATCTGATTTAGAAATCAGTCCCAAGATGATGAT 933
OY 1072 GTTAATATACACCCACACAGAAAGAGTTCCTTTGATTCAGAGCGTATTTGAA 1131
Db 934 GTTAATGTGACACCCACACAGATGATGATCTCTGACGAGAGAGCATCTCTGAG 993
OY 1132 ACAATTAAGATGCTATGAGAGAAACTGATGATTTCTATTAACACAGATATTCOA 1191
Db 994 CGGGTGCAGCAGCATGAGAGCAAGCTCCTGGGCTCCATTCCTCCAGATGACTTC 1053
OY 1192 ACTAGGCATT 1202
Db 1054 ACCCAGACTTT 1064

RESULT 10
AX234588 2484 bp DNA linear PAT 11-SEP-2001
LOCUS AX234588
DEFINITION Sequence 10 from Patent WO0162945.
ACCESSION AX234588
VERSION AX234588.1 GI:15593571
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Nicolaiides, N.C., Sass, P.M., Grasso, L., Vogelstein, B. and
Kinzler, K.W.
TITLE Methods for generating hypermutable yeast
JOURNAL Patent: WO 0162945-A 10 30-AUG-2001;
The Johns Hopkins University (US); Nicolaiides, Nicholas, C. (US);
Sass, Philip, M. (US); Grasso, Luigi (US); Vogelstein, Bert (US);
Kinzler, Kenneth (US)
FEATURES
source 1. 2484
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 720 a 535 c 594 g 635 t
ORIGIN

Query Match 12.8%; Score 319.8; DB 6; Length 2484;
Best Local Similarity 57.7%; Pred. No. 1e-54;
Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;

OY 172 ATCCGAGGTTGGAGAGTGGTGTGAACCCGATCGGGCGGGGAGGTATCCAGCGG 231
Db 43 ATTCGGCGCTGGAGAGACAGTGTGAACCCGATCGGGCGGGGAGGTATCCAGCGG 102
OY 232 CCGTGTGGCGGTGAAGAGAGTGTATCCAGAACACCTTCGAGCGGCGCCGCCGCGC 291
Db 103 CAGCTAATGCTATCAAGAGATGATGAGAACTGTTAGTCAAAATTCACAAATATT 162

QY 292 TCCGTCGGGTGAAGAGCGGTGCTCAAGCTATCCAGTCTCCATGACGGCCATGCG 351
 DB 163 CAAGTATTGTTAAAGAGGAGCGCTGAGATGATTCAGATCCAGCAATGACCGCGG 222
 QY 352 ATGAGGTTGAGATGTTGGCAATATTTGGCAAGGATCTCCCTCAAGTTATCGCA 411
 DB 223 ATGAGGAAAGAAATCTGGATATTGATGAAAGGTTCTACTGTAATGCACTGCGTC 282
 QY 412 TACGAGATCTGCGAGACCAATTAATGATGGGTTGACGAGGAGGCTTGGCTAGTAC 471
 DB 283 TTGAGGATTTAGCAGTATTTCTACCTATGCTTCGAGAGTGGGCTTTGGCAGATGA 342
 QY 472 ACTTATGTTGGCATGTTACCGTCGACAGCATTAACAGAGGCCAATTCACGGCTAC 531
 DB 343 AGCCATGTTGGCTGTTCTTACTATTAACAGCAAAACAGCTGATGAAAGTGTGATACGA 402
 QY 532 GTTCTTACAGATGTTGTAATGAGAAATGACCTTAAGCTTGGCTGGCGTGAAGA 591
 DB 403 GCAAGTTACTCAGATGAGAAAGCTGACCCCTCTTAACCATGTTGCTGCAATCAAGG 462
 QY 592 ACTCAAGTCATGTTGTAATTAATTTTACACATGTTAGCCCGCAGAAAGATGTCAG 651
 DB 463 ACCGATGTCAGCGTGGAGACCTTTTTCACATAGCCACGAGAGAAAGCTTTAAA 522
 QY 652 AACTCAATGATGATCAACCAAGATCGTAGCTTATCACTGCTGTTGCACTGCATCAC 711
 DB 523 AATCCAGTCAAGAAATATGAGAAATTTTGGAAAGTTGTTGGCAGTTCAGTACACAT 582
 QY 712 ATCAAGTTCCTTCTCTTCCAGAAAGATGAGCCCAATGACCAATGTTTCATAGTCA 771
 DB 583 GCAGGATTTAGTTCTCAGTTAAATAAACAAGAGAGACAGTACTGATGTTAGACACTA 642
 QY 772 AGTACATCTCAAGGTTAGATGCTATCAGAGAGAGTCTGCTGCTTCTGCTTGTGAT 831
 DB 643 CCCAATGCCCAACCGTGGAGCAATATTCGCTCATCTTTGGAATGCTGTAGTCGAGAA 702
 QY 832 CTCATGAAATTAAGTTTCATATGAGATGCTGACAGATTCATCTTCAAGATGATGT 891
 DB 703 CTGATGAAATTTGAGATGCTGAGATGAAACCTTACCTTCAAAATGATGATGT 753
 QY 892 TACATGCAATGCAATTAATGAGCAAGAAAGATTAATGATGATCTTTTCAATAATGAT 951
 DB 754 TACATGCAATGCAATTAATGAGCAAGAAAGATGATGATCTTTTCAATAATGATGAT 813
 QY 952 AGGCTGTAGACTGATCTGCTTGAAGAGACTTATGATGATGATGATGATGATGATG 1011
 DB 814 CGTCTGTAGACTGATCTGCTTGAAGAGACTTATGATGATGATGATGATGATGATG 873
 QY 1012 CTTCAAGATCAACACCTTTCATATGATGATGATGATGATGATGATGATGATGAT 1071
 DB 874 CCCAAAGACACACACCTTTCATATGATGATGATGATGATGATGATGATGATGAT 933
 QY 1072 GTTAATATATACCAACCAAGAAAGATGATGATGATGATGATGATGATGATGATG 1131
 DB 934 GTTAATATATACCAACCAAGAAAGATGATGATGATGATGATGATGATGATGATG 993
 QY 1132 ACATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1191
 DB 994 CGGCTGACAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1053
 QY 1192 ACTCAGGCAAT 1202
 DB 1054 ACCCAGACTTT 1064

RESULT 11
 HS007343
 LOCUS HS007343 2484 bp mRNA linear PRI 29-NOV-1995
 DEFINITION Human DNA mismatch repair protein homolog (hMLH1) mRNA, complete cds.
 ACCESSION U07343
 VERSION U07343.1 GI:463988
 KEYWORDS

SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2484)
 AUTHORS Bromner, C.E., Baker, S.M., Morrison, P.T., Warren, G., Smith, L.G., Lescoe, M.K., Kane, M., Earlbino, C., Liford, J., Linblom, A., Tannergard, P., Boilag, R.J., Godwin, A.R., Ward, D.C., Nordenskjold, M., Fisher, R., Kolodner, R.D., Ward, Liskay, R.M. Mutation in the DNA mismatch repair gene homolog hMLH1 is associated with hereditary non-polyposis colon cancer
 JOURNAL Nature 368 (6468), 258-261 (1994)
 MEDLINE 94195398
 PUBMED 8145827
 REFERENCE 2 (bases 1 to 2484)
 AUTHORS Morrison, P.T.
 TITLE Direct Submission
 JOURNAL Submitted (02-MAR-1994) Paul T. Morrison, Molecular Biology Core Facility, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 002115, USA

FEATURES
 source
 1..2484
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /map="3p22-23"
 1..2484
 /gene="hMLH1"
 22..2292
 /gene="hMLH1"
 /note="homolog of S. cerevisiae PMS1 (Swiss-Prot Accession Number P14242), S. cerevisiae hMLH1 (Genbank Accession Number U07187), E. coli MUTH (Swiss-Prot Accession Number P23367). Salmonella typhimurium MUTH (Swiss-Prot Accession Number P14161) and Streptococcus pneumoniae (Swiss-Prot Accession Number P14160)."
 /codon_start=1
 /product="hMLH1"
 /protein_id="AAC50285.1"
 /db_xref="GI:463988"
 /translation="MSFVAGVIRLDETVVNRVIAAGEVIOVRPAKIKEMENCLDAS
 TSIOVLKREGKLIQIDNGVIRKEDLDIVERTTSLOFEDLASTYFNGE
 ALASISHVAHVTITTTADGKCAVRSYSGKAPKAPAGNOQIIVEDVFNIA
 TRKALKNSPEVEGKILEVGRYSVHAGSFSKRGEGVADVIRLIPNASTVDNRA
 IFGNASREIEIGCEDKTLAFKNGVYSNASKVCKIFLFIINRLVSTSLRAK
 ERYVAAYLRKRPFLYSLIEISPOVDVAVHPTKHEVHRLHESTLIEROCHESKL
 LGSNSRMVETQTLILAGRSEGMKSTSLSSSTSGSDVYVYHNAVRTDSREK
 LDFLQPLSKPLSPQALVTEKTISSGRARQDEMELEPAPAEVAKNOSLED
 TTKGISEMSEKRPSTSNPKRRHRESDVEMVEDSRKEVTACTPRRRIINTSVLS
 LOEINIEGHEVREMLHNSFVGCNPQWALQHOTKLTILNTLSTLSELEFYILY
 DFANFGLRLSEPAFLDLAMLALDSPEGWTEDGPKGLAYIVLELKKRAEMLAD
 VSELEIDEGENLGLPLIDNVYVPELGLPIFLRLATEVKNMDEKCPESISKCAM
 FYIRKQIYSESTLSGQSEVSGSIPNSKMTVEHIVYKALSHILPRHFTEDGNI
 LQLANLPDIKVERC"

BASE COUNT 720 a 535 c 594 g 635 t
 ORIGIN

Query Match 12.8%; Score 319.8; DB 9; Length 2484;
 Best Local Similarity 57.7%; Pred. No. 1e-54;
 Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;

QY 172 ATCCGAGGTTGAGAGTGGGTGAGAACCGATCGCGGGGGGAGAGTATCCAGCG 231
 DB 43 ATTGGGGGCTGGACGAGACAGTGTGTAACCGCATCCGGGGGGAAGTATATCCAGCG 102
 QY 232 CCGTGTGGCGGTGAAGAGCTCATCGAAGAACGCTCGACGCTGCGCCCTCAGCGTC 291
 DB 103 CCACATATGCTATCAAGAAGATGATGAGAACTGTTAGTCAAAATACACAATAT 162
 QY 292 TCCGTCGGGTGAAGAGCGGTGATGATGATGATGATGATGATGATGATGATG 351
 DB 163 CAAGTATTGTTAAAGAGGAGCGCTGAGATGATTCAGATCAAGCAATGAGCAGCGG 222

352 ATCAGGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 411
 223 ATCAGGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 282
 412 TACGAGATCTGCAGACCAATTAATGATGAGGAGGAGGAGGAGGAGGAGGAG 471
 283 TTGAGAGATTTGAGGAGATTTCTACATATGAGGAGGAGGAGGAGGAGGAG 342
 472 ACTATGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 531
 343 ACCGATGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 402
 532 GTTCTTACAGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 591
 403 GCAAGTATCTGCAGACCAATTAATGATGAGGAGGAGGAGGAGGAGGAGGAG 462
 592 ACTCAGTATCTGCAGACCAATTAATGATGAGGAGGAGGAGGAGGAGGAGGAG 651
 463 ACCGATGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 522
 652 AACTCAGTATCTGCAGACCAATTAATGATGAGGAGGAGGAGGAGGAGGAGGAG 711
 523 ATCAGGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 582
 712 ATCAGGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 771
 583 GCAAGTATCTGCAGACCAATTAATGATGAGGAGGAGGAGGAGGAGGAGGAG 642
 772 ATCAGGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 831
 643 ACCGATGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 702
 832 CTGATGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 891
 703 CTGATGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 753
 892 TACATGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 951
 754 TACATGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 813
 952 AGGTTGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 1011
 814 CGTGTGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 873
 1012 CCTCAGTATCTGCAGACCAATTAATGATGAGGAGGAGGAGGAGGAGGAGGAG 1071
 874 CCTCAGTATCTGCAGACCAATTAATGATGAGGAGGAGGAGGAGGAGGAGGAG 933
 1072 GTTATGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 1131
 934 GTTATGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 993
 1132 ACATGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 1191
 994 CGGTGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 1053
 1192 ACTCAGTATCTGCAGACCAATTAATGATGAGGAGGAGGAGGAGGAGGAGGAG 1202
 1054 ACCGATGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 1064

RESULT 12 2530 bp mRNA linear PRI 12-JUL-2001
 LOCUS BC006850
 DEFINITION Homo sapiens, mult. (E. coli) homolog 1 (colon cancer, nonpolyposis type 2), clone MGC:5172 IMAGE:3451538, mRNA, complete cds.
 ACCESSION BC006850
 VERSION BC006850.1 GI:13905125
 KEYWORDS MGC.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2530)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (27-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 CONTACT Contact: MGC help desk
 EMAIL Email: cgabs-remail.nih.gov
 TISSUE Tissue Procurement: ATCC
 CDNA CDNA Library Preparation: Life Technologies, Inc.
 DNA DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 SEQUENC Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
 CENTER Center code: BCM-HGSC
 WEB Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 CONTACT Contact: villalona@bcm.tmc.edu
 VILLALON Villalon, D.K., Luna, R.A., Hale, S.M., Huliy, S., Lu, X., Garcia, A.M., Hollway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.
 CLONE Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 SERIES Series: IRAC Plate: 3 Row: 1 Column: 24
 THIS This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA 91: 466461.
 LOCATION Location/Qualifiers
 1. 2530
 /organism="Homo sapiens"
 /db_xref="locusid:4292"
 /db_xref="taxon:9606"
 /clone="MGC:5172 IMAGE:3451538"
 /tissue_type="Placenta, choriocarcinoma"
 /clone_id="NHL_MGC_10"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 51. 2321
 /codon_start=1
 /product="mult. (E. coli) homolog 1 (colon cancer, nonpolyposis type 2)"
 /protein_id="AA06850.1"
 /db_xref="gi:13905126"
 /translation="MSFVAGYIRLDEFTVNRKAGVRIORPANAIEKMEIENDLAKS TSIOYVIEGGLKTIQIIONDNGTIRKEDDIVERFTTSKLOSPEDASISGFRE ALASISHVAHTTTTKADKCAKRAVSQKTKAPKPCAGNGOITVEDLFYNA TRRLAKLPSEIEKILIEVGRSVNAGISFVKKGEIYADYRTIRPASTYDNRS IFVNAVSRLEIEICEDKTLAFKRNYSIVANVSFKCEIPLFTNRLVSTSLRKAI ETVVAAAYLPKTHPEPLYSLEISPDNDVYVHPKRVHFLHRESILEROVHIESKL LGSNSRMYFTQITLPGIAGSGEMVSTSLTSSSGSDKRYAHOVARTDSRECK LDATFLOPLSKESLOPQIVTEDKIDISGEARODDEMELETPAPAVAKNOSLED TTKGTSSEKRGFTSNPRKRHRSDVEVEDSREKEMTAETPRRIINTSVLS LOEELINEGHVILAEMLHNHSFVCNVPQNALAHOKRIYILNTKISELEFOILY DPAFNGVLRSEPAFLDPAALADSPESGTEEDREGIAETIVLEPKKEMILAD YFSEIDEGNMIGLPLINDYVPLDELPLFLRLATEVMMDEKCFSLSECKAM FYSIRKQYISESTLSGOOSVPSIPSMKMWVEHIVYALNRSHIILPPKHFEEDNT LQALNLPDYKFERC"

BASE COUNT 742 a 543 c 601 g 644 t
 ORIGIN
 Query Match 12.8%; Score 319.8; DB 9; Length 2530;
 Best Local Similarity 57.7%; Pred. No. 1e-54;
 Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;

172 ATCCGAGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 231
 72 ATCCGAGTTGGAGTATGGCAATATTTGGCAAGGACATCTACTCAAGTATCTCA 131
 232 CCCTGCTGCGGCTGAAGAGCTCATCGAAGAGCTGCGGCTGCGGCTGCGGCTGCGG 291
 132 CCAGCTAATGCTATCAAGAGATGAGTATGAGTATGAGTATGAGTATGAGTAT 191
 292 TCCGTTGCGTGAAGAGGCTGCTCAAGCTCATCAGTATGAGTATGAGTATGAGT 351

D	b		192	CAGAGATTGTTAAAGAGGAGGCCCTGAAGTTCATGCATCCAGACAAATGGCACCGGG	251
O	y		352	ATCAGGTTTGAGGATTTGGSCAATATTTGGCGAAAAGCAACTACTCAAGTTATCTGCA	411
D	b		252	ATCAGGAAGAAGATGTGGATATTGTATGTGAAGAGTTCACTACTAGTAAGTCCAGTCC	311
O	y		412	TACGAGATCTGCAGACCATAAATCGATGGGGTTTCAGAGGGAGGCTTTGGCTAGTAG	471
D	b		312	TTTGAGATTTTAGCCAGATTTTCTACGATATGGCTTTTCAGGTGAGGCTTTGGCCAGCATA	371
O	y		472	ACTTATGTGGCCATGTTACCGTGACAACGATACAGAAAGCCAAATTGCAGGCTACAGG	531
D	b		372	AGCCATGGGCTCATGTTACTATTACCAAGAAACAGCTGATGGAAGGTGTCATACAGA	431
O	y		532	GTTTTCTACAGAGATGGTGTATGGAGAAATGAGCCTTAAGCCTTTGGGCTGGCGTGAAGA	591
D	b		432	GCMAATTCTCAGATGAGAAACTGAAAGCCCCCTCTTAACCACTGGTGCTGGCAATCAGGG	491
O	y		592	ACTCAAGTCATGGTTGAAAAATCTATTTTTACAACATGGTAGCCCCGAGAAAAACATTGCA	651
D	b		492	ACCCAGATCAGCGGTGGAGAGACCTTTTATCAACATAGCACAGAGAAAGACCTTTAAAA	551
O	y		652	AATCTCAATGATGATACCCCAGAGTCTGTAAGCTTCATCAGTCGGTTTGCAATCATCAC	711
D	b		552	AATCCAACTGGAAGATATGAGGAAATTTTGGAAAGTGTGGAGATATTCATACACAAAT	611
O	y		712	ATCAACGTTTACCTCTCTTTCAGAAAGCATGAGGCCAATAGACAGATGTCATAGTGA	771
D	b		612	GCAGGCATTAATTTCTCAGTTATAAAAACAGAGAGACAGTAGTCGATGTTTAGACACTA	671
O	y		772	AGTACATCCTCAAGGTTTGATGATGCTATCAGAGAGTGTCTATGGGGCTTCTGCTGCTGAT	831
D	b		672	CCCAATGCTCTAACCGGTGAGCAATTTGTGCTCCATCTTTGGAAATGCTGTTAGTGAGAA	731
O	y		832	CTCACTGAATAAAGGTTTCATATGAGATGCTGCAGATTCATCACTTCAAGATGATGT	891
D	b		732	CTGANTGAATTTGATGT-----GAGATTAAGCCCTAGCCTTCCAAATGAAATGTT	782
O	y		892	TACATCTCAAAATGCAAAATTAATGTGGCAAGAAGATTAACATGATTTCTTTCATTAATGAT	951
D	b		783	TACATATCTCAATGCAAACTACTACATGAAAGTGCATCTTCTTACTCTTCATCAACAT	842
O	y		952	AGGCTTGAGACTGATGAGCTTTGAAAAGAGCTATGTAATTTGTACTCTCCAACTTG	1011
D	b		843	CGTGTGTGAGATCAACTCTCTGTGAAAAGGCATATGAAACAGTGTATGACACCTATTGG	902
O	y		1012	CCTCAAGATCCAAACCTTTCATATACATATGTCATACATCTCCATATCAGAACAGCTGAT	1071
D	b		903	CCCAAAACACACACACCACTCTCTGTAAGCTTATGAAATATGATGATGATGATGATGAT	962
O	y		1072	GTTAATTAACACCCACCAAGAGAGGTTAGCCTTTTGAATCAAGAGCGTATTAATGAA	1131
D	b		963	GTTAATGTGACACCCACCAAGAGCATGAAATTGCTCTGACAGAGAGAGCATCTCGAG	1022
O	y		1132	ACAATTAAGAAATGCTATTGAGGAAAAACTGATGATTTCTAATACAAACAGATATTCCA	1191
D	b		1023	CGGCTGCACGACGACATGAGACAGCAAGCTCTGGGCTCOAATTCCTCCAGGATGTAATTC	1082
O	y		1192	ACTCAGGCATT 1202	
D	b		1083	ACCCAGACTTT 1093	
<hr/>					
RESULT 13					
LOCUS AY069160					
DEFINITION Drosophila melanogaster GH18717 full length cDNA.					
ACCESSION AY069160					
VERSION AY069160.1 GI:17861655					
KEYWORDS FLI_CDNA;					
SOURCE Drosophila melanogaster.					
ORGANISM Drosophila melanogaster					
2068 bp mRNA linear INV 17-DEC-2001					

OY 279 GCGCTCCAGCGCTCCGTTGGGGTGAAGAGCGGTGGCTCAAGCTCATCAGTCTCCGA 338
 DB 129 ATCCGCCACATCCAGGTGACAGGTGAGCGGTGGCTGAGAGCTTCAATTCAGGA 188
 OY 339 TGAGGGCCATGGCATCAGGATTTGAGCATTTTGGCAATATTTGGCAAAAGGCAATACCTC 398
 DB 189 CAATGGCAGGGGATACGACGCGAGGATCTGGCCATCTGCGAGGCGGTTCACACCTC 248
 OY 399 AAGTATTCGATACGAGGATCTGACAGCCATAAATGATGGGCTCAGAGGAGGCG 458
 DB 249 CAGGTCAGCCCTTCGAGGATTTGTCCAGATACCAATTTGGCTCAGAGGAGGCG 308
 OY 459 TTTGGCTATGACTTATGCTTATGCTTACGCTGACCAAGATACAGAGGCGCAAT 518
 DB 309 CTGGGACACATTAAGTATGCTGCGCCTTAAGCATCCAGACAGAGGCGCTAGAGAA 368
 OY 519 GCAGGCTACAGGCTTCTTACAGAGATGCTTAATGAGAAATGAGCCCTTGGCG 578
 DB 369 GTGGCGCTACAGGCGCAGCTACGCGGATGCAAGCTCCAGGCTCAGCCGAGCGGTG 428
 OY 579 TGCGGTGAAGAGAACTCAAGTATGTTGAAATCTATTTTCAACATGCTAGCCGCA 638
 DB 429 CGGAACCAAGGACAGCATTTGCTATGAGGATCTGTTCTACAACTGCCCAAGAGG 488
 OY 639 GAAACATTCAGCAACTCCAAATGATGACCAACCAAGATCGTAGACTTCATCAGTGT 698
 DB 489 GCAGGACTGCTGCGCGCGAGAGTTCAGAGATTTCCGAGGTTCTGGCTAGTA 548
 OY 699 TGCAGTCAATCAGATCAGTATCTTCTTTCAGAGAAAGCATGAGCCATAGAGCAGA 758
 DB 549 CGGTGTCACAAATCCAAAGGTGGGCTTCACTCCGCAAGGAGAGCGCCCACTCC 608
 OY 759 TGTTCATATGCAAGTATCTTCAAGTATGATGCTATCAGAGTGTATGGGGCTTC 818
 DB 609 CCGCGGACACAGGAGTCCGATCCGATCCGAAACATCCGATCATTTACGGGCGAG 668
 OY 819 TGTGCTGCTGATCTATGAAATTAAGTTTCAATATGAGATGCTGCAAGTCAATCT 878
 DB 669 AATTTCCAGAGACTGCTGAGTTTATGCAACCGCATGAGGT-----TACAAAT 719
 OY 879 CAAGATGATGTTTACATCTCAATGCAATTTTGGCAAGAAAGATTACATGATTC 938
 DB 720 CGAGGCGAGTGTCTGATCTGATCTGATTAATCTGCGCAAAAGTGTAGAGTCT 779
 OY 939 TTTTCATTAATGATAGGCTTGTAGCTGATCTTGTGAAGAGATTTGATTTGTA 998
 DB 780 GTTCATTAACAGGACTGTGTGAGTCTGACAGCTTCAGAGACTCCGTGACATCGATTT 839
 OY 999 CTCTGCAACATTTGCTTACAGCATCCAAACCTTTCAATATACATGCTTCATCTCC 1058
 DB 840 TGGCACAATACCTGCGCGGTGGCCACATCTTGTGATATATAGCTGACGCTGCCGC 899
 OY 1059 AGAAGCAGTGTGATTAATATACACCCCAACCAAGAGAGCTTACGCTTTGATCAAGA 1118
 DB 900 ACAAACTTGTGAGCTGATTCATCCCAAGCATGAGAGCTGACATTTCTTCAACGGA 959
 OY 1119 GCGTATTAATGAACATTAAGAAATGCTATTGAGAGAAACTGATGATTTCAATACAC 1178
 DB 960 AGAGATTTGTGATTAAGCAAGCAGGTTGAAGCCCGGACTCTAGCAGCAATGCTCAC 1019
 OY 1179 CAGGATTTTCCAACTCAGGCTTAATACCTATATCAGGATTTGCTACCTAACCCACAAA 1238
 DB 1020 GAGAGCTTTTACAAACAGCTAAGGCTACAGAGGCCCGGACTGATGAATGAACCCAGTT 1079
 OY 1239 GGAATAGGTTTCTAGGCGCAGTATGGTCTGGAACAAATTTCTCAAAATTTCTGTGAG 1298
 DB 1080 GCGTGAATTAAGCAAGCAGCTTT-----ATCCAGAGAGATGTTGCGGTA 1123
 OY 1299 CCAATATGTTGAGACAGATCCCAATTCATCTGAGAGATTTGCAACACTTACTGGCAGG 1358
 DB 1124 CTGACTCCAGAGCAAAAGCTGATTAAGTTTGGGCGCTCTGTCAAAAGTACTCAG 1183

OY 1359 GC-AATCTCAAACTTTGAAAGAAATTTGATCTGTATCTGTGAAGAAATGTTGTAAGAT 1417
 DB 1184 GCGTATCTCTCAAGTCTCTCCAGGAAAGCTGTCTGCGCAGAAAGAGATTTGCGGTTA 1243
 OY 1418 CAAGGAAACCAAAAGATGCTGTATTTGTTCAGACCGCTCATGAGCTCTGTGGAAA 1477
 DB 1244 CAGGCGGCAAAATTCAGAGAACTGCTGTGAGCACCGCTGTAGACATAGAGAGCGG 1303
 OY 1478 TAGATTTAGCTTCATCTCAGGCTTTTGGACATTTGCAAGAACTGCATATATGTTGAC 1537
 DB 1304 TGAGCGGCACTGAGGCTTCAGCTGAGAGCAGCTCAAAACTTGTCTATGTGGGCT 1363
 OY 1538 TTGCGATGAAGCTTTGCTTGTATACACACAATACCCTTATACCTGTAAATGCG 1597
 DB 1364 GTGTGATGAGAGCGAGCCCTGTTCCAGCATGAAGACCGCTTATATGTGCAACAC 1423
 OY 1598 TAAATATAGTAAGAACTTATATGACCAAGAGCTTTGTGCGCTTTGGAACTTCATG 1657
 DB 1424 GGTGCTTAGAGGAATCTTCTATACGCGGATGATTTACGATTTCAAACTGCTCGG 1483
 OY 1658 CTATTCAGCTCAGTGAACACAGCTCCACTTCAGAGATTTGCTGTATGACACTGAAAGCG 1717
 DB 1484 AATTCACAAATAGTCCGCAATTTGCTCTCAGAGAGCTGTATCTTGTGCTGGAAAGCG 1543
 OY 1718 ATGAATGATGATGATGAAGAAAGATGATGAGAACTGAGATTTGCAAGATTAACACTG 1777
 DB 1544 AAGCGCGCGCTGACGCGCCAGAGCGGGGATTAAGGCTGAGACTGGCGAGCGTCTCGG 1603
 OY 1778 AGATCTAAAGAAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1837
 DB 1604 ACATCTTCTGAAAGAGCGCCCATCATGAGAGATTTTGGGCTGCGCATCTCGGAGG 1663
 OY 1838 ATGCAAAATGCAAGACTTCTGTTTGTATGAGACAGTACACCCCTGATATGAGCGTCT 1897
 DB 1664 ATGAATGCTGAATCTCTGCAAGTTGCTGACACAGCATGAGACTGCTGTGCGCAC 1723
 OY 1898 TTCCAGATTTTCTTGGCTTTAGAAATGATGATGATGATGATGATGATGATGATGAT 1957
 DB 1724 TGGCTTATCTGCTGCGCTTGGCCAGGAGTGTGAGCTGGAGCAGAGACAGATGCT 1783
 OY 1958 TCAGAACAGT 1967
 DB 1784 TCGAGACAT 1793

RESULT 14
 BC021815 2548 bp mRNA linear ROD 07-AUG-2002
 LOCUS Mus musculus, clone MGC:19289 IMAGE:4020489, mRNA, complete cds.
 DEFINITION BC021815
 ACCESSION BC021815.1 GI:18255307
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 2548)
 STRAUSBERG, R.
 Direct Submission
 Submitted (18-JAN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

REMARK
 COMMENT

Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
 Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
 Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/MLN at: <http://Image.lim.gov>
 Series: IRK. Plate: 24 Row: b Column: 13
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, Genomescan gene prediction, similarity but not identity
 to protein.

FEATURES

Source

Location/Qualifiers

1. 2548
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="CECH 11"
 /clone="MGC:19289 IMAGE:4020489"
 /tissue_type="Mammary tumor metastasized to lung."
 /library="MGI model. Expression driven by an MMTV-LTR
 enhancer."
 /clone_11b="NCI CGAP_Lu30"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6".
 39..2321
 /codon_start=1
 /product="unknown (protein for MGC:19289)."
 /protein_id="AAH21815.1"
 /db_xref="GI:18255308"
 /translation="MAYVAGYIRLDEYVNRIRAGVYIQRANAIKEMICLDAKS
 TNIOYVKEGKLIQIDNGIGIRKEDLDIVCEFTTSKLTQFEDLISISTYGRGE
 ALASISHVAHYITTKADPKACARASDGLQAPPRPCAGNOGTLTVDELFI
 TRRAKNPSEYEGKILEVGRYSIHNGISISVKGCTVSDVRLTNATVDELNRS
 IEGNAVRELLEVEGCEDETLAKFNKNGYSINAKSVKICIFLFIHRLVESALSKAI
 ETVYAYLIPKTHPLYSILSISPNONVNVPPHVEVHLHEESIIQRNOHIESKL
 LGSNSRMTFTOTLPLGLAPSGEAPRTTVASSSTGSGDKYATVQAVRSDOK
 LDAFLQVSLVSPQDPAPVPGARTGSGSPERARREDEMLAPAEVASENLE
 RESLMESTDIOKAPTSPPSSRRRHEDSDVEVNAKEMTAPRRIINTL
 SVLSIOEIISEKCHETLREMLNHSFVGCVPOMLAHQIYLYLNTTKISEEFYQ
 ILIDFANGVRLSEPAFLDLMLALDSPSGTIEDDGPREGLAETVIEVLEKRAE
 MLADYSEVIDEENLIGLPIILDSVPLREGLPFIILRIAEVMMDEKKEFESLSK
 ECAMFYSTRKOYIIEESTLSGQSDMPSTSKPMKMYEHIITKAFRSHLPKHFTF
 DGNVLQNLNLDLIVFERC"

CDS

BASE COUNT 709 a 599 c 651 g 589 t
 ORIGIN

Query Match 12.0%; Score 300.6; DB 10; Length 2548;
 Best Local Similarity 56.5%; Pred. No. 8.2e-51;
 Matches 583; Conservative 0; Mismatches 439; Indels 9; Gaps 1;

172 ATCCGAGGTTGAGAGATCGGTGTGTAACCGCATCGCGGGGAGGTGATCCAGCG 231
 60 ATTGCGGCTCTGGACAGACGCGTGTGAACCCGATAGCGGGGAGGATCAGCG 119
 232 CGGTGCGGCGGTGAGAGCTCATGGAAGCGCTGAGCGCTGGCGCTCCAGCGT 291
 120 CCGGCCAAATGCTATCAAGAGATGATAGAAACTGTTAGATGCAAAATCTACAATATT 179
 292 TCCGTGGGTGAAGAGCGGCTCAAGCTCATCCAGTCCCATGACGGCCATGGC 351
 180 CAAGTGTGTTAGAGAGGTGCGCTGAAGCTATTCATCAACAAATGCGCACTGGA 239
 352 ATCAGATTGAGAGATTGGCAATATTGTGCGAAAGCATCTACTCAAAAGTATCTGCA 411
 240 ATCAGAGAGAGATCTGATATTTGTGTGAGAGAGTTCACTACGAGTAACTGCAACT 299
 412 TACGAGATCTGACAGCACTAAATCGATGGGTTTCAAGAGGAGCGCTTGGCTACTAGT 471
 300 TTGAGAGATTGAGCGATTTCTACCTATNGGCTTCTGCTGAGAGCATTTGGCAACGATA 359
 472 ACTTATGTTGGCAGATGTTACCGTGACAAAGATTAACAGAGGCAATTCACGCGTACAG 531
 360 ACCCATGTGGCCATGCTATTAACMAAAGAGCTGATGGGAATGTGGGTACAGA 419

532 GTTTCCTACAGAGATGTTGTAATGAGAAATGAGACCTTAAGCCTTGCCTGCGGTGAAGA 591
 420 GCAAGTTACTAGATGAGAAAGCTGCAAGCCCTCTTAACCCCTGTCGACAGCAACGAGG 479
 592 ACTCAAGTATGTTTAAATCTATTTCACACATGATGACCCCGCAAGAAACATTCGAC 651
 480 ACCCTATGAGGTGGAAGACCTTTTTCACATATATCAACAGGAGGAAAGCTTTAA 539
 652 AACTCCATGATGACATACCCCAAGATGTAAGTCTATCATGACAGTGGTGGAGTCATCAC 711
 540 ATCCAGATGAGATGATGAGAAATTTTGAAGTGTGGACAGGATTCATACACAA 599
 712 ATCAAGTTACCTCTCTTGCACAGAAAGCATGAGCCCAATAGACAGATGTTCTAGTGA 771
 600 TCAGCATTAGTATCTCAGTTAAATAAACAAGGTGACAGATATCTGATGACAGACAG 659
 772 AGTACATCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 831
 660 CCCAATGCCCAACCGGTGCAACATTCCTCCATCTTGGAAATGCGGTACTGACAGAA 719
 832 CTCATGAAATTAAGGTTTCATATGAGATGCTGACATTCATCTCAAGATGATGATGAT 891
 720 CTCATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 770
 892 TACATCTCAATGCAATATATGTCGCAAGAGATGATGATGATGATGATGATGATGATGAT 951
 771 TATATATGCAATGCAATATATGTCGCAAGAGATGATGATGATGATGATGATGATGATGAT 830
 952 AGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1011
 831 CGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 890
 1012 CCGCAAGATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1071
 891 CCGCAAGATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 950
 1072 GTTATATACACCAACCAAGAAAGATGATGATGATGATGATGATGATGATGATGATGAT 1131
 951 GTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1010
 1132 ACATATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1191
 1011 CGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1070
 1192 ACTGAGCAT 1202
 1071 ACCGAGACTT.1081

RESULT 15
 AF250844
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 2283)
 AF250844
 AF250844.1 GI:7595953
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 2 (bases 1 to 2283)
 Kumaran, M. and Rao, M.R.S.
 Cloning of the cDNA of the Mult. homolog, MLH1 from mouse testis
 Kumaran, M. and Rao, M.R.S.
 Direct Submission
 Submitted (30-MAR-2000) Department of Biochemistry, Indian
 Institute of Science, CNR Rao Circle, Bangalore, Karnataka 560 012,
 India
 FEATURES
 source
 Location/Qualifiers
 1..2283

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 4, 2003, 18:33:37 ; Search time 626 Seconds
(without alignments)
8997.201 Million cell updates/sec

Title: US-09-954-950-1

Perfect score: 2501
Sequence: 1 cggcagcgagatttgcagtc.....aaaaaaaaaaaaaaaaa 2501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_101002:*

- 1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2501	100.0	2501	24	AAD36728
2	926.2	37.0	2381	24	AAD36729
3	319.8	12.8	2484	16	AAO90814
4	319.8	12.8	2484	16	AAO90815
5	319.8	12.8	2484	22	AAH75044
6	319.8	12.8	2484	22	AAH75047
7	319.8	12.8	2484	24	AAH39200
8	319.8	12.8	2484	24	ABK86091
9	319.8	12.8	2525	16	AAO97525

10	319.8	12.8	2543	22	AAH34941
11	315.2	12.6	1995	23	ABL17901
12	259.4	10.4	2515	23	AAH91395
13	249.8	10.0	4052	23	ABL17900
14	243.8	9.7	2577	16	AAO90941
15	212.8	8.5	3218	22	AAH75039
16	212.8	8.5	3218	22	AAH75039
17	138.8	5.5	3716	23	ABL04226
18	138.8	5.5	3716	23	ABL04226
19	132.6	4.9	1902	23	AAH54342
20	113.6	4.5	1890	23	AAH53227
21	113.6	4.5	1830121	17	AAH42063
22	102	4.1	549	21	AAH95307
23	99	4.0	2019	21	AAH95307
24	97.8	3.9	375	24	AAH21989
25	97.8	3.9	14273	19	AAH52173
26	97.6	3.9	1950	23	AAH55932
27	94.4	3.8	1950	23	AAH55471
28	91	3.6	2869	23	ABL25906
29	88.6	3.5	48275	21	AAH81501
30	88.6	3.5	349980	21	AAH21610
31	84.6	3.4	837096	21	AAH81489
32	84.4	3.4	2355589	24	AAH90521
33	84	3.4	5695	20	AAH20636
34	83.4	3.3	2133	23	AAH53161
35	81.4	3.3	17087	20	AAH12984
36	81	3.2	624	18	AAH74542
37	79.8	3.2	2007	23	AAH50300
38	79.8	3.2	2010	23	AAH51625
39	76.6	3.1	111309	20	AAH54599
40	75.6	3.0	1909	24	AAH20250
41	73.6	2.9	3063	16	AAH74245
42	73.6	2.9	3063	16	AAH74245
43	73.6	2.9	3063	22	AAH75042
44	73.6	2.9	3063	22	AAH75042
45	73.6	2.9	3063	24	AAH39198

ALIGNMENTS

RESULT 1	AAH36728	standard; cDNA; 2501 BP.
XX	AAH36728	
AC	AAH36728	
XX	21-AUG-2002	(first entry)
XX		
XX		
DE	Rice MLH1 cDNA.	
XX		
KW	Rice; MLH1: gene mutation; mismatch repair system; transposon tagging; tissue-specific promoter; herbicidal safener; male sterility; detection; pollen formation; base pair mismatch; gene; ss.	
KW		
XX		
OS	Oryza sativa.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	121..2295
FT		/tag= "Rice MLH1 protein"
PN	WO200224890-A2.	
PD	28-MAR-2002.	
XX		
XX	18-SEP-2001; 2001WO-US29088.	
PF		
XX		
PR	18-SEP-2000; 2000US-233124P.	
XX		
PA	(PION-) PIONEER HI-BRED INT INC.	
XX		
PI	Mahajan PB;	

Human colon cancer
Drosophila melanog
DNA encoding novel
Drosophila melanog
Murine wild type M
Nucleotide sequenc
Yeast MLH1 cDNA.
Drosophila melanog
Pseudomonas aerugi
Haemophilus influe
Haemophilus influe
Cat flea head and
Streptococcus poly
Human ORX polynuc
Streptococcus polyn
Streptococcus pneu
Streptococcus pneu
Drosophila melanog
N. meningitidis pa
Neisseria meningit
N. meningitidis pa
Genomic sequence o
Polynucleotide seq
Enterococcus faeca
Enterococcus faeca
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Borrelia burgdorfe
Bacillus lichenifor
Human DNA repair p
Nucleotide sequenc
Human PM51 cDNA.
Human mismatch rep
Human cDNA encodin


```
QY 1501 CTTTGGACATTTGTCAGAACTGACATATGTTGACTTGCCGATGAGAGCCTTGTCTTG 1560
DB 1501 CTTTGGACATTTGTCAGAACTGACATATGTTGACTTGCCGATGAGAGCCTTGTCTTG 1560
QY 1561 ATACAAACAAATACCCGCTTATACCTTGTAAATGTTGTAATTTAGTAAAGAACTTATG 1620
DB 1561 ATACAAACAAATACCCGCTTATACCTTGTAAATGTTGTAATTTAGTAAAGAACTTATG 1620
QY 1621 TACACAGAGCTTGTGCGCTTTGGAACTTCAATGCTTTCAGCTGACATGAAACAGCT 1680
DB 1621 TACACAGAGCTTGTGCGCTTTGGAACTTCAATGCTTTCAGCTGACATGAAACAGCT 1680
QY 1681 CCACCTCAGAGAGTTGCTGTGATGACACTGAAAGAGATGAATGATGATGAAAG 1740
DB 1681 CCACCTCAGAGAGTTGCTGTGATGACACTGAAAGAGATGAATGATGATGAAAG 1740
QY 1741 GATGATGAGAAATGAGAGATGTCAGAAAGTAAACACTGAGTACTTAAAGAAATGCTGAG 1800
DB 1741 GATGATGAGAAATGAGAGATGTCAGAAAGTAAACACTGAGTACTTAAAGAAATGCTGAG 1800
QY 1801 ATGATTAATGAGTACTTCTTATTCATATGATCAAGATGCAAAATGACAGAGCTTCT 1860
DB 1801 ATGATTAATGAGTACTTCTTATTCATATGATCAAGATGCAAAATGACAGAGCTTCT 1860
QY 1861 GTTGTACTGACAGCAGTACACCCCTGATGACCGCTTCCAGAAATTTGTGCTTTA 1920
DB 1861 GTTGTACTGACAGCAGTACACCCCTGATGACCGCTTCCAGAAATTTGTGCTTTA 1920
QY 1921 GGAATGAGTGTACTGAGATGACGAGAAAGATGCTTCAAGAACTGATGCTGCTGA 1980
DB 1921 GGAATGAGTGTACTGAGATGACGAGAAAGATGCTTCAAGAACTGATGCTGCTGA 1980
QY 1981 GGAACCTTATGACATCTTATCCCAATCTTCCAAATCTGGAATGGAGATTCAT 2040
DB 1981 GGAACCTTATGACATCTTATCCCAATCTTCCAAATCTGGAATGGAGATTCAT 2040
QY 2041 TTATACAGAAATATAGATTCATGCTGATGAACTGATGAACTGATGATATCA 2100
DB 2041 TTATACAGAAATATAGATTCATGCTGATGAACTGATGAACTGATGATATCA 2100
QY 2101 GATGAAATGACCTTGTATCAAGAACTTCTGCGGAGCAGACAGATGAGCCCAAGT 2160
DB 2101 GATGAAATGACCTTGTATCAAGAACTTCTGCGGAGCAGACAGATGAGCCCAAGT 2160
QY 2161 GAGTGGACCATTCAGCATGCTTGTTCATCCATGAGTGGACTTTTCTCAAGCCCGAG 2220
DB 2161 GAGTGGACCATTCAGCATGCTTGTTCATCCATGAGTGGACTTTTCTCAAGCCCGAG 2220
QY 2221 TCAATGGCAACAGATGAAAGCTTGTGACAGTGTCTTGGAGAACTCTACAGATT 2280
DB 2221 TCAATGGCAACAGATGAAAGCTTGTGACAGTGTCTTGGAGAACTCTACAGATT 2280
QY 2281 TTTGAAAGGCTTGTATCAATGAGAAATATAGAGCAGATGATGATGATGATG 2340
DB 2281 TTTGAAAGGCTTGTATCAATGAGAAATATAGAGCAGATGATGATGATGATGATG 2340
QY 2341 AGTGTGTTGAAATGCTTGTATTCACCGTATATGATGATGATGATGATGATG 2400
DB 2341 AGTGTGTTGAAATGCTTGTATTCACCGTATATGATGATGATGATGATGATGATG 2400
QY 2401 ACTGAGAAAGAAAGATGCTTGTATGAAATGAAAGTAAAGTAAAGTAAAGTAA 2460
DB 2401 ACTGAGAAAGAAAGATGCTTGTATGAAATGAAAGTAAAGTAAAGTAAAGTAA 2460
QY 2461 TATTCGATCAACCAAAAAAAAAAAAAAAAAAAAAA 2501
DB 2461 TATTCGATCAACCAAAAAAAAAAAAAAAAAAAAAA 2501
```

RESULT 2
AAD36729
ID AAD36729 standard; DNA: 2381 BP.

```
XX AC AAD36729;
XX XX 21-AUG-2002 (first entry)
DE Arabidopsis thaliana MLH1 DNA.
KM MLH1: gene mutation; mismatch repair system; transposon tagging;
KW tissue-specific promoter; herbicidal safener; male sterility; detection;
KW pollen formation; base pair mismatch; gene; ds.
OS Arabidopsis thaliana.
FH Key Location/Qualifiers
FT CDS 3..2216
FT /tag= a
FT /product= "Arabidopsis thaliana MLH1 protein"
PN WO200224890-A2.
PD 28-MAR-2002.
PF 18-SEP-2001; 2001MO-US29088.
PR 18-SEP-2000; 2000US-233124P.
PA (PION-) PIONEER HI-BRED INT INC.
PI Mahajan PB;
PI WPI: 2002-416283/44.
DR P-PSDB: AAE22979.
PT Novel rice MLH1 ortholog nucleic acid molecule for increasing
PT efficiency of targeted gene mutation or homologous recombination in a
PT plant and for generating plants with reversible male sterility
XX PS Disclosure: Page 86-88; 90pp; English.
XX CC The invention relates to isolated rice MLH1 orthologue nucleic acids. The
XX CC nucleic acid is useful for increasing the efficiency of targeted gene
XX CC mutation or homologous recombination in a plant, by transforming a plant
XX CC with expression cassette comprising the nucleic acid linked to a chemical
XX CC inducible promoter, transforming the plant with nucleic acid comprising
XX CC a sequence having a desired mutation or a sequence to be homologously
XX CC recombined, where the transformation occurs in the presence of chemical
XX CC compound capable of inducing the promoter and the plant's cellular
XX CC mismatch repair system is inhibited and selecting the transformed plants
XX CC that contained the mutation or homologously recombined nucleotide
XX CC sequence. The plant cellular mismatch repair system is inhibited through
XX CC the use of transposon tagging of an MLH1 gene, sense- and antisense-
XX CC suppression of an MLH1 gene, antibody binding to an MLH1 polypeptide or
XX CC its variant, and targeted mutagenesis of specific amino acid residues
XX CC encoded by an MLH1 gene. The nucleic acid is also useful for producing
XX CC reversible male sterility in a plant, by transforming a plant with an
XX CC expression cassette comprising a lexa DNA binding site embedded in a
XX CC tissue-specific promoter that drives expression in the plant operably
XX CC linked to the nucleic acid when expressed disrupts pollen formation or
XX CC function through inhibition of the plant's cellular mismatch repair
XX CC system, transforming the plant with a second expression cassette
XX CC comprising a nucleotide sequence encoding a lexa repressor protein
XX CC operably linked to a chemically-inducible promoter that drives expression
XX CC in the plant, and exposing the plant to a compound capable of inducing
XX CC the chemical-inducible promoter, to induce expression of lexa repressor
XX CC protein. The tissue-specific promoter is an anther-specific promoter
XX CC and the chemical-inducible promoter is a herbicidal safener. The
XX CC polypeptide encoded by the nucleic acid is useful for detecting,
XX CC locating, or removing a base pair mismatch (SNP). The present sequence
XX CC is Arabidopsis thaliana MLH1 DNA.
XX CC
XX CC Sequence 2381 BP; 721 A; 450 C; 548 G; 662 T; 0 other;
```

Query Match 37.0%; Score 926.2; DB 24; Length 2381;

Best Local Similarity 65.4%; Pred. No. 5.5e-210;
Matches 1408; Conservative 0; Mismatches 733; Indels 12; Gaps 3;

138 GGGAGCCCCCGGATCCGGAGGTGGAGAGTGGTGTGACCGCATCCGGCGGGG 217
139 GAGAGCCACCGAGATTCACCGTTAGAGATTCAGTGTGACCGGTATCCGAGTGTG 132
218 AGGTGATCCAGCGCGGTGCGGCGGTGAGAGAGTGTGATGAGAGAGTGTGAG 277
133 AGGTATCCAGCGGTGAGTGTGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 192
278 GCGCTCCAGCGGTGCGGCGGTGAGAGAGTGTGAGAGAGTGTGAGAG 337
193 ATTCAAGTTTCATAGCGGTGTTGATGAGAGAGTGTGAGAGAGTGTGAG 252
338 ATGAGCGCATGAGCATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 397
253 AGGAGCGTACGATGATGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 312
398 CAAAGTATCTGATGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 457
313 CGAAGTATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 372
458 CTGTGCTATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 517
373 CATTAATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 432
518 TGCAGCGTACGAGTGTGATGAGAGAGTGTGAGAGAGTGTGAGAGAG 577
433 TTGATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 492
578 CTGCGGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAG 637
493 CTGCTGTCAGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAG 552
638 AGAAGATGTCAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAG 697
553 GGAAGATGTCAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAG 612
698 TTGAGTGTGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 757
613 TGCTATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 672
758 ATGTCATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 817
673 ATGTCATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 732
818 CTGTCATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 877
733 CAGTTGCAAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 792
878 TCAAGATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 937
793 TTGATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 852
938 TTTCATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 997
853 TTTCATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 912
998 ACTGTCATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1057
913 ATGTCATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 972
1058 CAGATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1117
973 GGAAGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1032
1118 AGCTATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1177
1033 AATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1092
1178 CAGATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1237

1093 CTAGAGGTTTCAAGAGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1152
1238 AGGTATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1297
1153 GTGAT---TCTCAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1209
1298 GCCAATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1357
1210 ACAAATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1269
1358 GCGATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1414
1270 CCAAGCCCAAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1329
1415 GATCAGAGAGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1474
1350 GCGAAGAGAGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1389
1475 AATATGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1534
1390 GAGTTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1449
1535 GACTTCCGATGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1594
1450 GATGAGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1509
1595 TGTATATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1654
1510 TGTGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1569
1655 ATGTCATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1714
1570 ACGCATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1629
1715 ACGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1768
1630 AGGAGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1689
1769 TAAACATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1828
1690 TGAATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1749
1829 TGTATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1888
1750 TGTATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1809
1889 TGTATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1948
1810 TGTATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1869
1949 AAGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 2008
1870 AGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1929
2009 TGTATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 2068
1930 TTTTGCAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 1989
2069 CTGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 2128
1990 AGAAGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 2049
2129 TTTGCGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 2188
2050 TGTATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 2109
2189 CATCATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 2248
2110 CGTCATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 2169
2249 AGGTGATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 2301
2170 AGTATGATGATGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAG 2222


```
RESULT 3
AA090814
ID AA090814 standard; cDNA; 2484 BP.
XX
XX AA090814;
XX
XX 04-MAR-1996 (first entry)
XX
XX Human wild type MLH1 gene, a Mutl. homologue.
XX
XX hMLH1; wild type; Mutl. homologue; cancer diagnosis; mismatch repair;
XX tumour; susceptibility; mutation detection; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 22..2292
XX FT /*tag= a
XX FT primer_bind 118..135
XX FT /*tag= b
XX FT primer_bind 313..329
XX FT /*tag= c
XX
XX W09516793-1.
XX
XX 22-JUN-1995.
XX
XX 16-DEC-1994; 94MO-US14746.
XX
XX 09-DEC-1994; 94US-0352902.
XX 17-DEC-1993; 93US-0168877.
XX 08-MAR-1994; 94US-0209521.
XX
XX (DAND ) DANA FARBER CANCER INST INC.
XX (UOR-) UNTV OREGON HEALTH SCI.
XX
XX Baker SM, Bollag RJ, Bronner CE, Kolodner RD, Liskay RM;
XX
XX WPI: 1995-231583/30.
XX P-PSDB; AAR75785.
XX
XX Determ. of a mutation in a mutl. homologue or gene prod. in a tissue
XX PT - used to diagnose cancer susceptibility, and to identify and
XX PT classify a DNA mismatch-repair-defective tumour
XX
XX Claim 33; Fig 3; 16pp; English.
XX
XX AA090814 encodes AAR75785 the wild type hMLH1, a Mutl. homologue. A
XX mutation in a hMLH1 or hPMS1 nucleic acid isolated from a subject,
XX CC can be detected by comparing it with an analogous segment of the
XX CC above wild type allele. This method can be used to diagnose cancer
XX CC susceptibility, or to identify and classify a DNA mismatch-repair
XX CC defective tumour.
XX
XX Sequence 2484 BP; 720 A; 535 C; 594 G; 635 T; 0 other;
XX
XX Query Match 12.8%; Score 319.8; DB 16; Length 2484;
XX Best Local Similarity 57.7%; Pred. No. 9.3e-66;
XX Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;
```

```
QY 352 ATCAGGTTTGAGGATTTGGCAATATTGTGCGAAAGGCATACCTCAAGTTATGTGCA 411
DB 223 ATCAGGAAAGAGATCTGATATTGTATGTGAAGGTTCACTAGTAAACGCGATCC 282
QY 412 TACGAGATGTGCAGACCAATAAATGATGGGTTTCAGAGGAGGCTTTGGTATGATG 471
DB 283 TTTGAGGATTTTACAGTATTTCTTACCTTATGCGCTTTCAGAGTGAAGGCTTGGCAGCAT 342
QY 472 ACTTATGTTGGCCATGTATACGTCAGACAGATTAACAGAGCCCAATTGCAAGGCTACAGG 531
DB 343 AGCCATGTGCTCTATGTTATTTATTTACAGAGAAAGCTGATGAAAGTGTGATACAGA 402
QY 532 GTTTCCTTACAGAGATGTTGTAATGAGAAATGAGCTTAAGCCTTACGCTGCGTGGAAGA 591
DB 403 GCAAGTTACTACAGATGGAATAAAGTGAAGCCCTCCTTAACATGCTGTGCAATCAAGG 462
QY 592 ACTCAAGTCATGTTGTAATAATCTATTTTTACAAATGATAGCCCGCAAGAAACATTGCA 651
DB 463 ACCCAATCACAGGTGAGGACCTTTTTCACATAGCCACGAGAGAAAAGCTTTTAAA 532
QY 652 AACTCCATGATGATACCCCAAGATGCTAGCTTTCATGCTGAGTGGTTCAGTCCATCAC 711
DB 523 AATCCAGTCAAGATTAATGAGAAATTTTGGAAAGTTGTGCGAGTATTTAGTACACAT 582
QY 712 ATCAAGTTACCTTCTCTTGCAGAAAGATGAGCCAAATAGACAGATTTCAATAGTCA 771
DB 583 GCAGGATTTAGTTCTTCAGTTAAATAAACAAGAGAGACATAGTATGTTAGACACTA 642
QY 772 AGTACATCTCAAGGTTAGATGCTATGACAGATGTCTATGAGGCTTGTGCTGTGAT 831
DB 643 CCCAATGCTCCCAACCGTGCAAAATTCCTCCATCTTGGAAATGCTGTAGTCAGAA 702
QY 832 CTCATGAAATTAAGTTTCATATGAGGATGCTGACATTCATCTCAAGATGATGAT 891
DB 703 CTGATAGAAATTTGATGT-----GAGGATTAACCTTAGCTTCAAAATGATGT 753
QY 892 TACATCTCAATCAATTAATGTTGGCAAGAAAGATTAATGATTTCTTTCATTAATGAT 951
DB 754 TACATATCCAAATGCAAACTCTCAGTGAAGAAGTCAATCTTCTTCACTTCACTCAACAT 813
QY 952 AGGCTGTACAGTCTACTGCTTTGAAAAGAGCTAATTTGTTGTTACTGCAACATTTG 1011
DB 814 CGTCTGTATGAATCACTTCTTGAGAAAAGCCATGAAGAGATGATGAGCCTTAATTTG 873
QY 1012 CCTCAACATTCACCAACCTTCATATACATGTCATATCATCTTCATAGAGAACGTGAT 1071
DB 874 CCCAAAACACACACCCATTCCTGTACTTCAATTAATCACTCCCGAATGTGGAT 933
QY 1072 GTTAATATACCCCAACCAAGAAAGAGTTAGCCTTTTGAATCAAGCGTATTTATGAA 1131
DB 934 GTTAATGTGACCCCAAGAAAGATGAAGTTCACTCTGACAGAGAGACATCCTTGAG 993
QY 1132 ACATATGAATATGCTATTGAGAAAAGCTGATGATTTCAATACACAGATATTTCCA 1191
DB 994 CGGATGACAGACAGACATCGAGAGCAAGCTCTGCGCTCAATTCCTCAGAGATGATTC 1053
QY 1192 ACTCAGGCAAT 1202
DB 1054 ACCCAGACTTT 1064
```

```
RESULT 4
AA094015
ID AA094015 standard; cDNA; 2484 BP.
XX
XX AA094015;
XX
XX 15-JAN-1996 (first entry)
XX
XX Human mismatch repair pathway gene, MLH1.
XX
XX Mismatch repair; MSH2; primer; identification; defect; alteration;
```

KW cancer; tumour; vaccine; ds.
 XX OS Homo sapiens.
 XX key Location/Qualifiers
 FT CDS 21..3243
 FT /tag="a
 FT /product="human_mismatch_repair_protein_MLH1
 XX PN W0514085-A2.
 XX PD 26-MAY-1995.
 XX PF 17-NOV-1994; 94WO-US13385.
 XX PR 13-JUN-1994; 94US-0259310.
 PR 17-NOV-1993; 93US-0154792.
 PR 07-DEC-1993; 93US-0163449.
 XX PA (UYVE-) UNIV VERMONT & STATE AGRIC COLLEGE.
 PA (DAND) DANA FARBER CANCER INST.
 XX PI Fishel R, Kolodner RD, Reenan RAG;
 XX WPI: 1995-200377/26.
 DR F-PSDB; AAR76071.
 XX PT Determining alteration in human mismatch repair pathways - used in
 PT the diagnosis, prognosis and therapy of cancers and in screening
 PT assays
 XX PS Example 13; Page 237-239; 256pp; English.
 CC AAQ4015 is the human mismatch repair pathway gene MLH1. Defects or
 CC alterations in such a gene result in the accumulation of unstable
 CC repeated DNA sequences, a feature of a number of different cancers.
 CC The identification of a defect in the mismatch repair pathway can
 CC be diagnostic of a predisposition to cancer and prognostic for a
 CC particular mammalian cancer e.g colorectal, ovarian, endometrial
 CC (uterine), renal, bladder, skin, rectal and bowel. The nucleotide
 CC sequences and polypeptides of the hMSH2 gene may also be used for
 CC therapy and in vaccines.
 XX SQ Sequence 2484 BP; 720 A; 535 C; 594 G; 635 T; 0 other;
 Query Match 12.8%; Score 319.8; DB 16; Length 2484;
 Best Local Similarity 57.7%; Pred. No. 9.3e-66;
 Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;

QY 532 GTTCTTACAGAGATGCTAAATGAGAAATGAGCCTTAACCTTGGCGGTGAAGA 591
 DB 403 GCAAGTTACTACAGATGGAAGAAAGCCCTTAACCATGTGTGGCAATCAAGGG 462
 QY 592 ACTCAAGTCATGCTGAAATCTATTTTACACATGTGTGCCCAAGAAACATTGGAG 651
 DB 463 ACCAGATTCAGGCTGAGAGACCTTTTTCACACATGACAGAGAGAAAGACCTTAA 522
 QY 652 AACTCCAGATGATGATACCCCAAGATGATGATGATGATGATGATGATGATGATG 711
 DB 523 AATCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 582
 QY 712 ATCAAGCTTACCTTCTCTTTCAGAGAGATGAGCCATGAGCAGATGATGATGATG 771
 DB 583 GCAAGCATTTGTTTCTCAGTTAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 642
 QY 772 AGTACATCTTCAAGGTTAGATGCTATCAGAGAGAGAGAGAGAGAGAGAGAGAG 831
 DB 643 CCCAATGCTTCAACCGGAGCAATATTCCTCATCTTGGAAATGCTGTAGTCAGAA 702
 QY 832 CTCATGAAATGAAAGCTTTCATATGAGAGATGCTGAGATTCATTCAGATGATGAT 891
 DB 703 CTGATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 753
 QY 892 TACATCTCAATGCAATTTATGTGCAAGAGATTTCAATGATGATGATGATGATG 951
 DB 754 TACATATCCCAATGCAATTTATGTGCAAGAGATTTCAATGATGATGATGATG 813
 QY 952 AGGCTTGTAGAGCTGATGCTTTGAAAGAGATGATGATGATGATGATGATGATG 1011
 DB 814 GCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 873
 QY 1012 CCTCAAGCATCAACCTTTCATATGATGATGATGATGATGATGATGATGATGAT 1071
 DB 874 CCCAATGCTTCAACCGGAGCAATATTCCTCATCTTGGAAATGCTGTAGTCAGAA 933
 QY 1072 GTTATATACACCCCAACCAAGAGATTTGATGATGATGATGATGATGATGATG 1131
 DB 934 GTTATGTGACCCCAACCAAGATGATGATGATGATGATGATGATGATGATGAT 993
 QY 1132 ACAATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1191
 DB 994 CGGTGTCAGAGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1053
 QY 1192 ACTCAGCAT 1202
 DB 1054 ACCCAGACTTT 1064
 RESUME 5
 AAH75044
 ID AAH75044 standard; CDNA; 2484 BP.
 XX AAH75044;
 29-OCT-2001 (first entry)
 DE Nucleotide sequence of human mismatch repair protein MLH1.
 KW MLH1; mismatch repair gene; MMR gene; hypermutable yeast; ss.
 XX Homo sapiens.
 OS
 FT key Location/Qualifiers
 FT CDS 22..2292
 FT /tag="a
 FT /product="MLH1"
 XX PN W0200162945-A1.
 XX PD 30-AUG-2001.
 XX PF 21-FEB-2001; 2001WO-US05447.

XX 23-FEB-2000; 2000US-0184336.
 XX (UYXO) UNIV JOHNS HOPKINS.
 XX (NICO/) NICOLAIDES N C.
 XX (SASS/) SASS P M.
 XX (GRAS/) GRASSO L.
 XX (VOCE/) VOGELSTEIN B.
 XX (KINZ/) KINZLER K W.
 XX Nicolaides NC, SASS PM, Grasso L, Vogelstein B, Kinzler KW;
 XX WPI: 2001-522820/57.
 XX P-PSDB: AAG63956.
 XX
 XX Making hypermutable yeast that exhibit novel selected output traits for
 XX commercial applications, comprises introducing polynucleotide
 XX containing dominant negative allele of mismatch repair gene -
 XX
 XX Disclosure: Page 44-45; 60pp; English.
 XX
 XX The present sequence encodes human MLH1. MLH1 is a mismatch repair
 XX (MMR) gene. The specification describes a method for making a
 XX hypermutable yeast, comprising introducing a polynucleotide containing
 XX a dominant negative allele of a mismatch repair (MMR) gene, into a
 XX yeast, whereby the cell becomes hypermutable. The method is useful
 XX to create desirable output traits for commercial applications, using
 XX dominant negative alleles of mismatch repair proteins. The hypermutable
 XX yeast is useful for production, biocatalysis, bioremediation and drug
 XX discovery. It is also useful in genetic screens for the direct selection
 XX of variant subclones that exhibit new output traits. The hypermutable
 XX yeast is also useful in the manufacturing industry for the generation
 XX of new biochemicals, for detoxifying noxious chemicals from by-products
 XX of manufacturing processes or those used as catalysts, for remediation
 XX of toxins present in the environment including polychlorobenzenes, heavy
 XX metals and other environmental hazards for which there is a need to
 XX remove them from the environment. The yeast is also useful for producing
 XX increased quantity or quality of protein or non-protein therapeutic
 XX molecule e.g., Penicillin G, Erythromycin and clavulanic acid, by
 XX biotransformation.
 XX
 XX Sequence 2484 BP; 720 A; 535 C; 594 G; 635 T; 0 other;
 XX
 XX Query Match 12.8%; Score 319.8; DB 22; Length 2484;
 XX Best Local Similarity 57.7%; Pred. No. 9.3e-66;
 XX Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;

DB 403 GCAAGTTACTCAGATGAGAAAGCCCTTCCTTAACCAAGTGTGGCAATCAAGG 462
 OY 592 ACTCAAGTCATGTTGTAAGAAATCTATTATTACACATGTTAGCCGCAAGAAACATTCAG 651
 DB 463 ACCCAATGACGCGTGGAGACCTTTTTCACATATGCCCCAGAGGAAAGCTTTTAA 522
 OY 652 AACTCAATGATGATCTACACCAAGATGCTGACCTTCAATGCTGGTTCAGTCATC 711
 DB 523 AATCCAGTGAGAAATATGAGAAATTTTGGAGTTGTTGGCAGTATTCAGTACCAAT 582
 OY 712 ATCAAGTTACCTCTCTTCCTTCAGAAAGATGAGCCCAATGACAGATGTTCAATGCA 771
 DB 583 GCAGCATTTAGTTTCTCAGTTTAAACACAGAGAGACAGTACGTGATTTAGACACTA 642
 OY 772 AGTACATCTCAAGGTTAGATGCTATCAGAGAGTGTATGAGGCTTCCTGCTTCGAT 831
 DB 643 CCCAATGCGTCACACCGTGGACATATTCGCTCATCTTTGGAATGCTGTACTCGAGAA 702
 OY 832 CTCATGAAATTAAGGTTTCATATGAGATGCTGCAGATTCAATCTTCAAGATGATGT 891
 DB 703 CTGATGAAATTTGATGT-----GAGATTAACCCCTACGCTTCAAAATGATGT 753
 OY 892 TACATCTCAATGCAATTTATGTCGCAAGAAAGATTAATGATTTCTTCAATATGAT 951
 DB 754 TACATATCCAAATGCAAACTGCTAGTGAAGATGCAATCTTCTTACTCTTCATCAACAT 813
 OY 952 AGGCTTGATGACATGCTGCTGTTGAAAGAGATTTGAAATTTGTTACTGCTGCAACAT 1011
 DB 814 CGCTGTGATGATCACTTCTTCTGAGAAAGCCATGAAACATGATGATGACCTATTG 873
 OY 1012 CTTCAAGCATCAACCTTTCATATGATGTCATATCATCTTCATCAGAACAGCTGAT 1071
 DB 874 CCAAAACACACACCATCTCTGATCTGATGTTGAAATGATGCTCCAGAAATGATGAT 933
 OY 1072 GTTAATATACACCAACCAAGAAAGATGATGCTTTGTAATGAAGCTATTATTGAA 1131
 DB 934 GTTAATGTCACCCCAACAAAGATGAAGTTCCTTCTGACAGAGAGACATCTGAG 993
 OY 1132 ACAATGAAGATCTATTGAGAAAGATGATGATTTCTATCAACACGATATTCCAA 1191
 DB 994 CGGCTGACAGACACATCGAGAGCAACCTCTGGCTCCAAATTCCTCCAGATGTACTTC 1053
 OY 1192 ACTCAGGCATT 1202
 DB 1054 ACCCAAGCTTT 1064
 XX
 XX RESULT 6
 XX AAH76367
 XX ID AAH76367 standard; cDNA; 2484 BP.
 XX
 XX AC AAH76367;
 XX
 XX AC
 XX
 XX DT 29-OCT-2001 (first entry)
 XX
 XX DE Human MLH1 cDNA.
 XX
 XX KW Hypermutable bacteria; mismatch repair gene; MMR gene; MutL; MutS; MutL;
 XX MutL; PMS2; MLH1; MMR; Biocatalysis; bioremediation; biochemical;
 XX drug discovery; detoxification; toxin; biotransformation; ss.
 XX
 XX OS Homo sapiens.
 XX
 XX FH Key Location/Qualifiers
 XX CDS 22..2292
 XX FT /tag= a
 XX FT /product= "MLH1"
 XX
 XX PD WO200159092-A2.
 XX
 XX PD 16-AUG-2001.
 XX
 XX PF 12-FEB-2001; 2001WO-US04339.

Disclosure; Page 18; 68pp; English.

Sequence 2484 BP; 720 A; 535 C; 594 G; 635 T; 0 other;

1

[illegible]

OY	699	TGAGTCATACACATCAAGCTTACCTCTCTTGCAAAAGCATGTGACCAATAGAGCA	758
Db	555	CGCTGTCCACAATCCAAAGGGGTGGGCTACTCTCCGCAAGCAGGAGACGGCAACCTGC	614
OY	759	TGTCATAGTGCAGATGACATCTCAAGGATAGATGCTATCAGAGTGTATGAGGGCTTC	818
Db	615	CCCTGGGACACACAGTGGCCAGTTCCCGATCCGAAAACCTCCGATCATTTACGGGGCAGC	674
OY	819	TGTGCTTGTGATCTCATVAGAAATTAAGTTTCATVATGAGATGCTGCAGATTCATCTT	878
Db	675	AATTTCCAGAGAGCTGCTGGAGCTTTAGCCACCGCATGAGT-----TTACAAGTT	725
OY	879	CAATATGGATGGTACATCTCAATTCGAATTTGTGGCAAGAAAGATTTACATATATCT	938
Db	726	CGAGCCGAGTGTGTGATCTCACTCAGGTAACTATTCGGCCAAAAGTGTAGATGCTCT	785
OY	939	TTTTATTAATGATAGGCTTTAGACTGTACTGCTTTGAAAAGAGCATTTGAATTTGTGTA	998
Db	786	GTTCATCAACGAGGACTGGTGGAGTCCACAGCTCTCAGAGCACTCGTGTGATCATTTA	845
OY	999	CTTGCAACATTTGCTTCACAGCATCAACCTTTCAATATACATGTCCATACATCTTCCATC	1058
Db	846	TGCCACATACCTGGCCCGGTGGCCACACATCTCTTGCATATATGACCTTCAGCTCCGCC	905
OY	1059	AGAACGTTGATGTTAATTTATACCAACCAAGAAAGAGTTAGCTTTGATATACAGA	1118
Db	906	ACAGACTTGGACCTGATGATTCATCTCCACCAAGCATVAGGTGCACCTTCTCTACACAGA	965
OY	1119	GCGTATTTATTAACCATTAAGAAATGCTATTGAGAAAAAGTATGAAATTTCAATACAC	1178
Db	966	AGAGTTGTGTGATAGCTTTAAGCAGCAGGTTGAAGCCCACTCTTAGCCAGCATGTCCAC	10235
OY	1179	CAGGATATTCCAAACTCAGGCATTAAACTTATCAGGAGTATGCTCAAGCTAACCCCAAAA	12388
Db	1026	GAGGACGTTTCAAAACAGCTAAGGCTACACAGAGGCCCGCACTGTGATGAACCCAGTT	10855
OY	1239	GGATTAAGTTTCTAGGCCAGTATGGTTGTGAAACAAAATCTCAAAAATTCCTGTGAG	12988
Db	1086	GGCTGATTAACCGCAAGCATTT-----ATCCAAAAGATAGTGGCTGA	11295
OY	1299	CCAATGGTGCAGACAGATCCAGCACAATCCATCTGGAAAGATTGCACACTACTGGCAGG	13585
Db	1130	CTGACTCCACAGACCAAAAGCTGATTAAGTTTGTGGCGCTCTGGTCAAAAATGACTAG	11885
OY	1359	GC-AATCTTCAATCTTGAAGAAAGAAATTTGATCTTATCTGTAAAGAAATGTTTGAAT	14175
Db	1190	GGCTATCTCTCAAGTCTCTCCAGAAACGCTGCTGTGCAGAAAGAGATTTTCGCTTA	12494
OY	1418	CAAGAGAAACCAAAAGATGTCGTGTATTTGTGCAGCCGCTATGACTGCTCTTGAGAAA	14777
Db	1250	CAGCGGCAAAAAAATVCAACGAGAGTGGCTTTGACGCTGTAGCAATGAGAGAGCCG	13095
OY	1478	TAGATVTCAGCTTCATCTCGCCCTTTTGACATTTGTCAAGAAATCCACATATGTTGAC	15375
Db	1310	TGGAGCGGCACTGCAGGCTTCAGCTGAGGAGCAGCTCAAAAATCTTGCTTATGTGGCT	13665
OY	1538	TTGCGGATGAAGCCTTGTGCTTTGATACACACAAATACCCGCTTATACCTTGAATATGG	15975
Db	1370	GTTGGATGAGAGCGCAGCCCTGTTCCAGATGAAGAGCGCCCTTTATATGTCAACACAC	14225
OY	1598	TAAATATTAGTAAAGAACTTATGTACCAAGACCTTTGTGCGTTTGGAACTTCAATG	16575
Db	1430	GGTGCTTATGAGGAGCACTGTCATACAGGGATTAACGATTTCAAACTCTGCGG	14895
OY	1658	CTATTCAGCTACAGAACCAAGCTCCACTTTCAGAGATTTGCTGGTATGGATCGCAAGACG	17175
Db	1490	AAATACAAATTAAGTCGCGCAATGTGCTCTCAAGAGAGTGTATCTTGTGCTGGAAACG	15495
OY	1718	ATCAATTGATGATGAGTGAAGAAAGATGATAGAAATCGAGATTCGAGATGAACACTG	17777
Db	1550	AAGCGCGCGCTGGAGCGCCCGGAAGAGCGGGATTAAGCTGAGCTGGCGGAGCGGCTGCGG	16095
OY	1778	AGATACTAAAGAAATGCTGAGTGTATTAATGATGACTTTTCTATTCATATGATCAAG	18375

OY 1361 AATCTCAATCTGTAAGAAATTTGATCTTGTATCTGTAAAGAAATGTTGTAAGATCA 1420
 DB 1803 TATCTCTCAGTTCCTCCAGAGAGGCTGCTGCTCCAGAGAGGTTTCTGCTTACAG 1744
 OY 1421 GGAGAAACCAAAAGATGCTGATTTGTCAAGCCGATAGAGCTCTTGTGMAATAG 1480
 DB 1743 CGGCAAAAATCAGCAAGATGCGTTTGACAGCGTCTAGACTATGAGAAAGCCGCTGG 1684
 OY 1481 AATTCAGCTTCATCTGCTGCTTTTGAACATTTGACAACTGCAATATGTTGACTTG 1540
 DB 1683 AGCGCAGTGCAGCTTACGTAGAGAGACAGCTCAAAAACCTTGCTATGCTGCTGTG 1624
 OY 1541 CGATGAAGCCCTTCTTGTATGACAAACAAATACCCGCTTATCTTGAATATGTGTA 1600
 DB 1623 TGGATGAGAGCGGAGCCCTGTTCCAGATGAAGAGCCCTTATATGTGCAACACAGCT 1564
 OY 1601 AATATGATGAAGACTTATGTATGACCAAGCTTTGCTGCTTTGGAACTTCAATGCTA 1660
 DB 1563 CGTTTNGTGAAGACTTCTTATGACGAGATTTAGAGATTTCAAACTGCTCGGAA 1504
 OY 1661 TTTACGCTCAGTGAACAGCTCCATTCAGAGATTTGCTGATGCTGCTGAAAGAGATG 1720
 DB 1503 TCACATATAGTCCGCTATGCTTCAAGAGAGCTGCTATCTTGTGCTGGAAGAGAG 1444
 OY 1721 AATGATGATGATGAAGAAAGATGATGAGAACTGGAATGTCAGAAATGAACTGAGA 1780
 DB 1443 CGCGCGCTGAGACGCCCGGAAGAGCGGAGATGAAGCTGAGCTGCGGAGAGCTGCGGAGCA 1384
 OY 1781 TACTAAGAAAGAAATGCTGATGATGATTAATGATGATCTTTCTATCATATGATCAAGATG 1840
 DB 1383 TTTCTTGTGAAGAGGCGCCATCATGAGATATTTGCGGCTGCGCATCTCGAGAGATG 1324
 OY 1841 GCAATATGCAAGACTTCTGTTGTATGATGACAGATGACCCCTGATGATGAGCCGCTTC 1900
 DB 1323 GAATGCTCAATCTCTGCCAGATTTCTGACACAGATGATGATGATGATGATGATGATG 1264
 OY 1901 CAGAAATTTGTTGCTTGTGAGAAATGATGATGATGATGATGATGATGATGATGATG 1960
 DB 1263 CTTGTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1204
 OY 1961 GAACACT 1967
 DB 1203 AGACATT 1197
 RESULT 14
 AAQ0941
 ID AAQ0941 standard; DNA; 2577 BP.
 XX
 AC AAQ0941;
 XX
 DT 05-MAR-1996 (first entry)
 XX
 DE Murine wild type MLH1 gene, a MutL homologue.
 XX
 KM mMLH1; wild type; MutL homologue; cancer diagnosis; mismatch repair;
 XX
 OS tumour; susceptibility; mutation detection; ss.
 XX
 Mus musculus.
 XX
 FH Key Location/Qualifiers
 FT 3..2188
 FT /tag- a
 FT /note- "The above CDS is putative and the START
 FT codon is absent.
 FT The reading frame has to be shifted 5' by
 FT one base at position 944 to decode to the
 FT amino acid."
 XX
 MO9516793-A1.
 XX
 PD 22-JUN-1995.

XX
 PF 16-DEC-1994; 94MO-US14746.
 XX
 PR 09-DEC-1994; 94US-0352902.
 PR 17-DEC-1993; 93US-0168877.
 PR 08-MAR-1994; 94US-0209521.
 XX
 PA (DAND) DANA FARBER CANCER INST INC.
 PA (UWOR-) UNIV OREGON HEALTH SCI.
 XX
 PI Baker SM, Bollag RJ, Bronner CE, Kolodner RD, Liskay RM;
 XX
 DR WPI: 1995-231583/30.
 DR E-PSDB; AAR75789.
 XX
 XX
 PT Determ. of a mutation in a mutL homologue or gene prod. in a tissue
 PT classed to diagnose cancer susceptibility, and to identify and
 PT classify a DNA mismatch-repair-defective tumour
 PS Example 2; Fig 12; 168pp; English.
 XX
 CC AAQ0941 encodes AAR75789 the predicted mMLH1, a MutL homologue. A
 CC mutation in a hMLH1 or hPMS1 nucleic acid isolated from a subject,
 CC can be detected by comparing it with an analogous segment of the
 CC claimed wild type mMLH1. This method can be used to diagnose cancer
 CC susceptibility, or to identify and classify a DNA mismatch-repair
 CC defective tumour.
 XX
 SQ Sequence 2577 BP; 705 A; 611 C; 644 G; 617 T; 0 other;
 XX
 Query Match 9.7%; Score 243.8; DB 16; Length 2577;
 Best Local Similarity 54.8%; Pred. No. 1.1e-47;
 Matches 532; Conservative 0; Mismatches 427; Indels 12; Gaps 2;
 OY 222 CGCTGCTGCGGCTGAAGAGACTCATGAGAAACAGCTGAGCGCTGCGCCCTCCAGCCTC 291
 DB 3 CGGCAATTCCTATCAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 62
 OY 292 TCCGTTGCGGTGAAGAGAGCGTGGCTCAAGCTCATCCAGTCTCCAGTACAGAGCCAGTGC 351
 DB 63 CAAGTGTGTTTAAAGAGAGTGGCTGAGAGCTTAATTCAGATTCAGAAAGAAATGCGAC 122
 OY 352 ATCAGTTTGAAGATTTGGCAATATTTGGCGAAAGCATCTACCTCAAGTTATCTCA 411
 DB 123 ATCAGAGAGAGAGATCTGATATTTGTGTGAGAGCTTCACTACAGATGAAGTGAAGCT 182
 OY 412 TACAGAGATCTGACAGACCAATTAATGATGGGTTTCAAGAGGAGGCTTGGCTAGTATG 471
 DB 183 TTTGAGAGTTTACAGATTTCTTCTACATATGCTTCTGTTGAGACATTTGGCAAGCATTA 242
 OY 472 ACTTATGTTGGCCATGTTTACCTGTCACAAAGATTAACAGAAAGCCATTTGACGGCTACAG 531
 DB 243 ACTCATGTGGCCCATGCTCATTTACCAACCAAAAGCTGATGGAATGTGCGTACAGA 302
 OY 532 GTTCTTACAGAGATGCTATGATGAGATGAGATGAGCTTAAGCTTGGCTGCGTGAAGA 591
 DB 303 GCAAGTTACTCAGATGGAAGAGCTGCAAGCCCTCTTAACCTGTGTCAGAGCAACAGGGC 362
 OY 592 ACTCAAGTCATGTTGAAATCTATTTTACACATGATGAGCCGCAAGAAACATTTGACG 651
 DB 363 ACCCTGATCAGGTGGAAGACCTTTTTCACATTAATTCACAGAGAGAGAGAGCTTTAAA 422
 OY 652 AACTCAATGATGATCAGCCCAAGATCTGATGATGATGATGATGATGATGATGATGATG 711
 DB 423 AATCCAAAGTGAAGAGTGAAGAAATTTTGAAGTGTGAGAGGATTTCAATACACAT 482
 OY 712 ATCAAGTACTTCTCTTGCAGAAAGATGAGCCATGAGCAATGAGATTTATAGTGA 771
 DB 483 TCGAGATAGTATCTCAATTAATAAAGAGTGAAGAGATGATGATGATGATGATGATGATG 542
 OY 772 AGTACATCTCAAGGTTTATGATGATGATGATGATGATGATGATGATGATGATGATG 831
 DB 543 CCATATGCCACAACCGTGAACAACATTCCTCATCTTTGGAAGAGGCTTGTAGTGAAGA 602


```

Db 923 TTCATATCAGCAAGTAGAAGATTAA---CTGGAAGCGTGAAGGAAAGGTGTGA 979
QY 902 ATGCAATTTATGTGCAAGAG---ATTACATCATCTTTTCATAATGATAGCCTTG 958
Db 980 ATTGAATTTTCATATCCAAAAGTCCATTTCTATTAATTTTTCATTAATATAGACTAG 1039
QY 959 TACACTGTACTGCTTGAAGAGCTATTTGATTTGTGTACTCTGCACATTGGCTCAG 1018
Db 1040 TGACATGTATCTTAAAGAGAGCTTTGACAGCGCTTACTCCAAATATCTGCCAAGG 1099
QY 1019 CATCAACCTTTCATATATCATATCATCTTCATCAGACAGCGTGAATTAATA 1078
Db 1100 GCTTCAGACCTTTTATTTATTTGGAAITGTATATAGATCCGGCGCTGTGATGTTACG 1159
QY 1079 TACACCCAGCAAGAGAGAGTTCAGCTTTTGAATCAGAGCGTATTTATGAAACATAA 1138
Db 1160 TTCACCCGACAAAGAGAGAGTTCGTTCCAGCCAGATGAGATCATAGAGAAATCG 1219
QY 1139 GAAATGCTATTTAGAGAAACTGATGATTTCTAATACACAGATATTCCAAACT 1194
Db 1220 CCAATCAATTGCAAGCGCAATATCTGCCATGATATCTTACCTTCAAGGCT 1275

```

Search completed: April 4, 2003, 20:04:56
 Job time : 662 secs

DR N-PSDB: AAD36728.
XX
XX Novel rice MHL ortholog nucleic acid molecule for increasing
PT efficiency of targeted gene mutation or homologous recombination in a
PT plant and for generating plants with reversible male sterility -
XX
XX Claim 7; Fig 2; 90pp; English.
DS

CC The invention relates to isolated rice MHI orthologue nucleic acids. The
CC nucleic acid is useful for increasing the efficiency of targeted gene
CC mutation or homologous recombination in a plant, by transforming a plant
CC with expression cassette comprising the nucleic acid linked to a chemical
CC inducible promoter, transforming the plant with nucleic acid comprising
CC a sequence having a desired mutation or a sequence to be homologously
CC recombined, where the transformation occurs in the presence of chemical
CC compound capable of inducing the promoter and the plant's cellular
CC mismatch repair system is inhibited and selecting the transformed plants
CC that contained the mutation or homologously recombined nucleotide
CC sequence. The plant cellular mismatch repair system is inhibited through
CC the use of transposon tagging of an MHI gene, sense- and antisense-
CC suppression of an MHI gene, antibody binding to an MHI polypeptide or
CC its variant, and targeted mutagenesis of specific amino acid residues
CC encoded by an MHI gene. The nucleic acid is also useful for producing
CC a reversible male sterility in a plant, by transforming a plant with an
CC expression cassette comprising a lexA DNA binding site embedded in a
CC tissue-specific promoter that drives expression in the plant operably
CC linked to the nucleic acid when expressed disrupts pollen formation or
CC function through inhibition of the plant's cellular mismatch repair
CC system, transforming the plant with a second expression cassette
CC comprising a nucleotide sequence encoding a lexA repressor protein
CC operably linked to a chemically-inducible promoter that drives expression
CC in the plant, and exposing the plant to a compound capable of inducing
CC the chemical-inducible promoter, to induce expression of lexA repressor
CC protein. The tissue-specific promoter is an anther-specific promoter
CC and the chemical-inducible promoter is a herbicidal safener. The
CC polypeptide encoded by the nucleic acid is useful for detecting,
CC locating, or removing a base pair mismatch (SNP). The present sequence
CC is rice MHI protein.

50 Sequence 724 AA;

Query Match	100.0%	Score 3709	DB 23	Length 724
Best Local Similarity	100.0%	Prod. No. 1.8e-297		
Matches 724	0	Mismatches 0	Indels 0	Gaps 0

Oy	1	MDESPRCCGACGEPRIRLRLEESVNRJAAGEYIQRPSSAVKELINSJLDAAGSSVSA	60
Dd	1	MDESPRCCGACGEPRIRLRLEESVNRJAAGEYIQRPSSAVKELINSJLDAAGSSVSA	60
Oy	61	VKDGGLKLIQVSDGHHGIFREDLAILCERHTTSKLSAYEDLQITKSGFGEALASMTYV	120
Dd	61	VKDGLKLIQVSDGHHGIFREDLAILCERHTTSKLSAYEDLQITKSGFGEALASMTYV	120
Oy	121	GHVYVTTTBEOLHGYSRYSDGVMENBPKCAAVKTOYVVELFPMVBARKKTLONSN	180
Dd	121	GHVYVTTTBEOLHGYSRYSDGVMENBPKCAAVKTOYVVELFPMVBARKKTLONSN	180
Oy	181	DDYPRKYDFEISRFVAVHINVTFSCKRGKANADVASHASTSRJDAISYVGAASYRDLIE	240
Dd	181	DDYPRKYDFEISRFVAVHINVTFSCKRGKANADVASHASTSRJDAISYVGAASYRDLIE	240
Oy	241	IKVSEYEDADADSIFKMDGYISNANTVAKKTIWILFINRLVDCTALKAIEFVYSATLPPQA	300
Dd	241	IKVSEYEDADADSIFKMDGYISNANTVAKKTIWILFINRLVDCTALKAIEFVYSATLPPQA	300
Oy	301	SKRPFYMSIHLPSEHVYVNIHPTKKEVSLNQEKLIIETIRALIEEKLMSNSTRIRIPOTQA	360
Dd	301	SKRPFYMSIHLPSEHVYVNIHPTKKEVSLNQEKLIIETIRALIEEKLMSNSTRIRIPOTQA	360
Oy	361	LNLSGIAOANOKOVASASMSGSKKSOIKPVSOVMFTDPDNPBSCGRJHTYWHGQSSULEK	420
Dd	361	LNLSGIAOANOKOVASASMSGSKKSOIKPVSOVMFTDPDNPBSCGRJHTYWHGQSSULEK	420

QY	421	KEDJVSANVNSRRNQDADJSSRHELLVEIDSSPHGLDLYVKCTYVGLADEAFAL	480
Db	421	KFDJVSANVNSRRNQDADJSSRHELLVEIDSSPHGLDLYVKCTYVGLADEAFAL	480
QY	481	IQHNTRLVAVVNIISKELMQALCFRGFNNAIQLSEPAPELJLALMKDELMSDEK	540
Db	481	IQHNTRLVAVVNIISKELMQALCFRGFNNAIQLSEPAPELJLALMKDELMSDEK	540
QY	541	DDEKLEIAEVNTEILKEAEMINEYFSIHIDQDGLTRLRPVLDQYTPMDRLPEFVLAL	600
Db	541	DDEKLEIAEVNTEILKEAEMINEYFSIHIDQDGLTRLRPVLDQYTPMDRLPEFVLAL	600
QY	601	GNDVYTWDDKECFERTVASAVGNFYALHPPIIPNPSGNGIHLKYKNRDSMADEHAENDLIS	660
Db	601	GNDVYTWDDKECFERTVASAVGNFYALHPPIIPNPSGNGIHLKYKNRDSMADEHAENDLIS	660
QY	661	DENDVDDGLLAFAARAAQAQREMTIQHVLFPSPMLFLPKPKSMADGCFYOVASLEKLYKI	720
Db	661	DENDVDDGLLAFAARAAQAQREMTIQHVLFPSPMLFLPKPKSMADGCFYOVASLEKLYKI	720
QY	721	FERC 724	720
Db	721	FERC 724	720

RESULT 2
AAE08710
ID: AAE08710 standard; Protein; 737 AA

DT 15-NOV-2001 (first entry)

DE Arabidopsis thaliana PMS2 protein homologue MLH

KW Hypermutable plant; dominant negative allele; mismatch repair gene;
MMR; cell line generation; PMS2; AtMLH.

OS Arabidopsis thaliana

PN 100 WO200161012-A1

PD 23-AUG-2001

PF 28-DEC-2000; 2000WO-US35397.

PR 18-FEB-2000; 2000US-0183333.

PA (NICO/) NICOLAIDES N C

PA: (SASS/) SASS P M.

PA (VOGE/) VOGELSTEIN B

PI Nicolaides NC, Grasso L, Sass PM, Kinzler K, Vogelstein B,

DR WPI; 2001-529913/58

Making hypermutable cell, useful for generating hypermutable plants, especially for crop plants with new traits

PT₃ polynucleotide comprising dominant negative allele of mismatch repair gene into plant cell

XX 5
PS 1 Example 1: Page 57-58: English

[illegible]

The method involves introducing into a plant cell a polynucleotide

The method is useful for generating hypermutable plants, new cell lines and plant materials which are particularly useful for screening.

important crops. The method is also useful for generating crop plants

CC-BY commercial use. The present sequence is *Arabidopsis thaliana* (At)

CC MLH protein. This sequence is a homologue of MMR protein, PMS2.

XX Sequence 737 AA;

Query Match 67.5%; Score 2505; DB 23; Length 737;

Best Local Similarity 66.4%; Pred. No. 5,9e-198;

Matches 482; Conservative 111; Mismatches 129; Indels 4; Gaps 3;

```

OY 2 DEPSRGGGCGAGPPRIRLEESVNRVIAAGEVITOPSSAVKELIENSIDAGASSVAV 61
DB 13 EESPAATTIVPREPKIQRLSESVNRVIAAGEVITOPSSAVKELVENSIDADSSISVAV 72
OY 62 KGGGLKLIQVSDGSHGIREDELAICERTTSLKSLAYEDLQITKSGFGEALASMTYVG 121
DB 73 KGGGLKLIQVSDGSHGIREDELAICERTTSLKSLAYEDLQITKSGFGEALASMTYVA 132
OY 122 HTVVTITTEGOLHGYRVSVDGVMEHPPKCAVKGTOVMENLFYNNVARKKTIONSD 181
DB 133 HTVVTITTEGOLHGYRVSVDGVMEHPPKCAVKGTOVMENLFYNNVARKKTIONSD 192
OY 182 DYKPIVDYFISRAVHIINTFSCRKGANRADVSHASSTSRDLAIRSYGASVARDLIEI 241
DB 193 DYKPIVDYFISRAVHIINTFSCRKGANRADVSHASSTSRDLAIRSYGASVARDLIEI 252
OY 242 KYVEDADSIKMDGYISNANYVAKKITMILFINDRLVDCALKAILEVYSATLIPQAS 301
DB 253 EYSSDSDSCCTFMEGFISNSNVAKKITLVIFINDRLVDCALKAILEVYSATLIPQAS 312
OY 302 KPIFVMSIHLPEHVDVNHPTKKEVSLNOERLIETINNAIEKLMNENTRIPTQAL 361
DB 313 KPIFVMSIHLPEHVDVNHPTKKEVSLNOERLIETINNAIEKLMNENTRIPTQAL 372
OY 362 NLGIAQANPQDKYSEASMSGTSOKIPVSQVNTDPRNPSGRILHTWHGSSMLEKK 421
DB 373 EYIQ-STLISQKSDSPVSOQKPSGOKQKVPKMKVRTDSDPAGRLAHATLPKPSGL 431
OY 422 FDLVS-VRRNVVSRNRQKAGDLSHHELVIEDSFHGLDIYKNCITYGIADAFAFL 480
DB 432 VSSLVSRSSVRQRNRPKFTALDSVQELIAGVDSCHGEMLETNYNCITYGADDFAL 491
OY 481 IOHNFRLYVNVVNIISKELTMOALCRFGNFNAIOLSEPAILOELLVMAIKDEL--MSD 538
DB 492 VOYNTHLIYANVNSKELTMOALCRFGNFNAIOLSEPAILOELLVMAIKDEL--MSD 551
OY 539 EKDEKLEIAEVNTELIKEMNINEYSIHIDODKLTRLPVLIDQYTPMDRLPEFVL 598
DB 552 TKDOLKERIAEMNTELIKEMNINEYSIHIDODKLTRLPVLIDQYTPMDRLPEFVL 611
OY 599 ALGNDVYMDDEKERTVAVSANGFYALHPILPNFSGGIIHLYKKNRDSMADEHAENDL 658
DB 612 CLGNDVYMDDEKERTVAVSANGFYALHPILPNFSGGIIHLYKKNRDSMADEHAENDL 671
OY 659 ISDENVDDELAEAAVAAVAREMTIQHLPFPMRLFKPPSMARDGTFVYVASLEKLY 718
DB 672 VMEDELDDLLSDAENAAVAREMTIQHLPFPMRLFKPPSMARDGTFVYVASLEKLY 731
OY 719 KIFERC 724
DB 732 KIFERC 737

```

RESULT 3

AAE22979

AAE22979;

21-AUG-2002 (first entry)

Arabidopsis thaliana MLH1 protein.

MLH1: gene mutation; mismatch repair system; transposon tagging;
tissue-specific promoter; herbicidal safener; male sterility; detection;

```

KW pollen formation; base pair mismatch.
XX Arabidopsis thaliana.
OS WO200224890-A2.
XX 28-MAR-2002.
XX 18-SEP-2001; 2001WO-US29088.
XX 18-SEP-2000; 2000US-233124P.
PR (PION-) PIONEER HI-BRED INT INC.
PA Mahajan PB;
PI WPI: 2002-416283/44.
DR N-PSDB: MAD36729.
XX Novel rice MLH1 ortholog nucleic acid molecule for increasing
PT efficiency of targeted gene mutation or homologous recombination in a
PT plant and for generating plants with reversible male sterility
XX Disclosure: Page 89-90; 90pp; English.
PS
XX The invention relates to isolated rice MLH1 ortholog nucleic acids. The
XX nucleic acid is useful for increasing the efficiency of targeted gene
XX mutation or homologous recombination in a plant, by transforming a plant
XX with expression cassette comprising the nucleic acid linked to a chemical
XX inducible promoter, transforming the plant with nucleic acid comprising
XX a sequence having a desired mutation or a sequence to be homologously
XX recombined, where the transformation occurs in the presence of chemical
XX compound capable of inducing the promoter and the plant's cellular
XX mismatch repair system is inhibited and selecting the transformed plants
XX that contained the mutation or homologously recombined nucleotide
XX sequence. The plant cellular mismatch repair system is inhibited through
XX the use of transposon tagging of an MLH1 gene, sense- and antisense-
XX suppression of an MLH1 gene, antibody binding to an MLH1 polypeptide or
XX its variant, and targeted mutagenesis of specific amino acid residues
XX encoded by an MLH1 gene. The nucleic acid is also useful for producing
XX reversible male sterility in a plant, by transforming a plant with an
XX expression cassette comprising a lex A DNA binding site embedded in a
XX tissue-specific promoter that drives expression in the plant operably
XX linked to the nucleic acid when expressed disrupts pollen formation or
XX function through inhibition of the plant's cellular mismatch repair
XX system, transforming the plant with a second expression cassette
XX comprising a nucleotide sequence encoding a lex A repressor protein
XX operably linked to a chemically-inducible promoter that drives expression
XX in the plant, and exposing the plant to a compound capable of inducing
XX the chemical-inducible promoter, to induce expression of lex A repressor
XX protein. The tissue-specific promoter is an anther-specific promoter
XX and the chemical-inducible promoter is a herbicidal safener. The
XX polypeptide encoded by the nucleic acid is useful for detecting,
XX localizing, or removing a base pair mismatch (SNP). The present sequence
XX is Arabidopsis thaliana MLH1 protein.
XX
XX Sequence 737 AA;
XX
XX Query Match 67.5%; Score 2505; DB 23; Length 737;
XX Best Local Similarity 66.4%; Pred. No. 5,9e-198;
XX Matches 482; Conservative 111; Mismatches 129; Indels 4; Gaps 3;

```

[illegible]

RESULT 4	
AAW09036	
ID	AAW09036 standard; protein: 752 AA.
XX	
XX	
AC	AAW09036;
XX	
DT	01-OCT-1997 (first entry)
XX	
DE	Mismatch repair protein, hPMS2.
XX	
KW	Human; mismatch repair protein; hMSH2; hMLH1; hPMS2; antibody; Ab;
KW	proliferation; anti-cancer therapy; neoplastic tissue; DNA replication;
KW	DNA repair; mutation; sporadic colorectal carcinoma; cancer; hNPCC;
XX	inherited; non-polyposis colorectal carcinoma.
XX	
OS	Homo sapiens.
XX	
PN	W09641192-A1.
XX	
PD	19-DEC-1996.
XX	
PF	07-JUN-1996; 96WO-US08957.
XX	
PR	07-JUN-1995; 95US-0480351.
XX	
PA	(UYJO) UNIV JOHNS HOPKINS.
XX	
PI	Burrell M, Hill DE, Kinzler KW, Vogelstein B;
XX	
XX	
DR	WPL; 1997-077287/07.
XX	

PT. Discriminating proliferating from non-proliferating cells in tissue
PT. - using antibodies specifically immuno-reactive with mismatch repair
XX. protein, esp. human MSH2
PS. Disclosure; Page 23-25; 37pp; English.
XX

The sequences given in AAW09034-36 represent the human mismatch repair proteins hMSH2, hMLH1 and hPMS2. In the method of the invention, these proteins were identified by reaction with an antibody (Ab) specific for them, thereby discriminating proliferating from non-proliferating cells. The method may be used for monitoring the effectiveness of anti-cancer therapy in neoplastic tissue, by comparing the amount of Ab- γ g complexes in the sample with an amount determined at an earlier time, in which a reduction in the amount indicates an effective therapy. The Ab are especially specifically immunoreactive with the MSH2 mismatch repair gene, which is 1 of at least 4 genes encoding proteins involved in the repair of mismatched nucleotides following DNA replication or repair. Mutations in the MSH2 gene contribute to the development of sporadic colorectal carcinoma, while germline MSH2 mutations are responsible for approx. 50% of inherited, non-polyposis colorectal carcinoma (HNPCC). Since MSH2 is ubiquitously expressed, development of other cancers are also susceptible to alterations in MSH2.

Sequence 752 AA;

Query Match	39.3%	Score 1457	DB 18	Length 752
Best Local Similarity	39.7%	Pred. No. 2.5e-111		
Matches 311	Conservative 146	Mismatches 212	Indels 114	Gaps 11

18 IRLEESVNRIAGEVIORPSSAVKELIENSLDAGASSVSVAVKDGLKLIQVSDDGHG 77

8 IRLDET VNR IAG E V I Q R P A N A I K E M I E N C L D A K S T S I Q V I V K E G L K L I Q I D N G T G 67

78 IRFEDLAICERHTTSKLSAYEDIQTIKSMGFGEALASMTYVGHVTTITEGOLHGYR 13

68 IRKEDLDIVCERFTTSKLQSFEDLASISTYGFGEALASISHVAHVTITTKTADGKCAZR 12

138 VSRDGVME NEPKPCAAVKGIQVMENLFYNNVARKKTLQNSND DYPKIVDFISREAVHH 19

120 ASDGALNAPFNCAGNQGILVEDLEYNIAIRKALKNPSEYKILEVGRYSVHN 18

100 120 140 160 180 200 220 240 260 280 300 320 340 360 380 400 420 440 460 480 500 520 540 560 580 600 620 640 660 680 700 720 740 760 780 800 820 840 860 880 900 920 940 960 980 1000

[illegible][illegible]

STATE OF NEW YORK
IN SENATE
JANUARY 11, 1911.
REPORT
OF THE
COMMISSIONERS OF THE LAND OFFICE
IN RESPONSE TO A RESOLUTION
PASSED BY THE SENATE
MAY 1, 1909.
ALBANY:
J. B. LIPPINCOTT & COMPANY, PRINTERS.
1911.

30E

376 WCEBVCCECECKTIVCQUMEDDNDNDNDY WWWWQCCSUY

364 TSTSSSTGSSDKVVAHOMITDSEB- :| : | | | : |
db :| : | | | | |

0Y 427 VRNVB-----SRRNOKDAGDIS----- 44

422 GRARODEEMLELPAPAEVAAKNOSLEGDTTKGTSEMSFEKRGPTSSNPBKRBHREDSVEM 48

445 ---SRHELLV-----EIDSSFHPGLDIVKNCTYVGLADEAFALJ 48

482 VEDDSRKEMTACTPRRIINTSVLSLOEEINEOGHEVLREMLHNHSFVGCVPNOMATA 54:

QY 482 QHNTRLVNVNISKELMYQALCREGNFNAIQLSEPAPIQELLVMALKDDELMSEKD 541

Db 542 QHQTLYLNTFKLSEELFYQILIYDFANFVRLSEPAFLFDLMLALDSPESGWTEED 60

542 DEKLEIAEVNTEILKENAEMINEYFSIHIDQDGKTLRPLVLDQYTPMDRLPEFVLALG 60

602 GPKEGLAEYIVEFLKKKAEMLADYFSLEIDEENLGLPLLDNYVPPLEGLPIFILRLA 663

QY 602 NDYTWDEKECFRTVASANGNYALHPILPNPNSGNGILYKKNRDSMADEHAENDLISD 661
 DB 662 TEVNMDEKECFESLSKECAMFYSI-----RKQYISE 693
 QY 662 ENDVDOELAEAAQAOREMTIOHVLPPSMRLFLKPKSMATDGTGFVOVASLEKTKIE 721
 DB 694 ESTLSGQ-0SEVPGSW---KMTVEHLYKALRSHILPKHFTEDGNITLDANLPDLKYF 749
 QY 722 ERC 724
 DB 750 ERC 752

RESULT 5

AAR75785
 ID AAR75785 standard; Protein; 756 AA.

AC AAR75785;

DT 04-MAR-1996 (first entry)

DE Human wild type MLH1, a Mutl homologue.

KM bMLH1: wild type: Mutl homologue; cancer diagnosis; mismatch repair:
 tumour; susceptibility; mutation detection.

OS Homo sapiens.

PN WO9516793-A1.

PD 22-JUN-1995.

PF 16-DEC-1994; 94WO-US14746.

PR 09-DEC-1994; 94US-0352902.

PR 17-DEC-1993; 93US-0168877.

PR 08-MAR-1994; 94US-0209521.

PA (DAND) DANA FARBER CANCER INST INC.

PA (UYOR-) UNIV OREGON HEALTH SCI.

PI Baker SM, Bollag RJ, Bronner CE, Kolodner RD, Liskay RM;

DR WPI: 1995-231583/30.

DR N-PSDB: AAO90814.

PS Claim 33; Fig 3; 168bp; English.

CC AAO90814 encodes AAR75785 the wild type bMLH1, a Mutl homologue. A

CC mutation in a bMLH1 or hPMSt nucleic acid isolated from a subject,

CC can be detected by comparing it with an analogous segment of the

CC above wild type allele. This method can be used to diagnose cancer

CC susceptibility, or to identify and classify a DNA mismatch-repair

CC defective tumour.

CC Sequence 756 AA:

Query Match 39.1%; Score 1452; DB 16; Length 756;

Best Local Similarity 39.6%; Pred. No. 6.5e-111;

Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;

DB 18 IRRLESYVNRITAGEVIOFPSAVKELIENSDDAASSVSVAVKGGKLLQVSDGHC 77
 8 IRRLEDTVNRITAGEVIOFPSAVKELIENSDDAASSVSVAVKGGKLLQVSDGHC 67
 78 IRRLEDTVNRITAGEVIOFPSAVKELIENSDDAASSVSVAVKGGKLLQVSDGHC 137
 68 IRRLEDTVNRITAGEVIOFPSAVKELIENSDDAASSVSVAVKGGKLLQVSDGHC 127

QY 138 VSTRGVMEENREKPCAAVGTQVMEENLEFNMARAKKTLQNSDDYPKIYDFTSRVAHH 197
 DB 128 ASYSDCKLAPKPCACAGNOGTQITVEDLFPNATRRKALKNPSEEGKILEVVGGRSVNH 187
 QY 198 INTFESCRHGANRADVHASTSSRLDAIRSVYGASVVDLIEIKSYEDAASITKMG 257
 DB 188 AGISFESVKKQGETADVRLPNASTVDNIRSTFGNVSRLLEIGGEDKTLA---FKMG 244
 QY 258 YISNANYVAKKITMILFINDRLVDCATLAKRAIEFVVSATLPQAKRPFIWASHLPESEHD 317
 DB 245 YISNANYSVKCTFLFIFHRLVVESTSLKRAIETVYAALPKRTHFPIVLSLEISQND 304
 QY 318 VNIHPYKKEVSLINOERIETIRNALIEKLMNSNTTRIFOTQALNLSGIAQAQNPQ--KDK 375
 DB 305 VVNHPTKHEVHFLHEESILERVQOHIESKLSGNSRMVFTQTL-LPGLAGPGEVWKST 363
 QY 376 VSEASMSGCTKOKIPVSOQVTRDPNRPNSGRLLTFYHGGSSNL-----KKKEDVLS 426
 DB 364 TSLTSSSTSGSSDKVYAHQVNRDTSRQ-KIDAFIQPLSKPLSOPQALVTEDKTDISS 421
 QY 427 VRNVVR-----SRNQKDAGLS----- 444
 DB 422 GRARQDDEMLLPAPAEVAAKNQSLGDTTKTSEMSEKRGPTSSNPKRHREDSDVEM 481
 QY 445 -----SRHELIV-----EIDSFHGLDLYKNCYVGLADEAPAL 481
 DB 482 VEDDSRKEMTAACPPRRRIINLTSLVSLQELNEQHEVLRMLAHNSFVGCYVPMQWAL 541
 QY 482 OHNTRLIVNVNISKELVQALCRPGNPNATOLSEPAIOLRLVMAKDDLMODEX 541
 DB 542 OHOTKLYLNTKTLSELEFYQIILYDFANGVLRLESPALFLAMALDSPSGWTEED 601
 QY 542 DEKLEIAEVNTEILKENAEMINEYFSIHIDODKLTRLPAVLDOYTPDMRLPEFVALG 601

QY 602 NDYTWDEKECFRTVASANGNYALHPILPNPNSGNGILYKKNRDSMADEHAENDLISD 661
 DB 662 TEVNMDEKECFESLSKECAMFYSI-----RKQYISE 693
 QY 662 ENDVDOELAEAAQAOREMTIOHVLPPSMRLFLKPKSMATDGTGFVOVASLEKTKIE 721
 DB 694 ESTLSGQ-0SEVPGSW---KMTVEHLYKALRSHILPKHFTEDGNITLDANLPDLKYF 749
 QY 722 ERC 724
 DB 750 ERC 752

RESULT 6

AAR76071
 ID AAR76071 standard; Protein; 756 AA.

AC AAR76071;

DT 15-JAN-1996 (first entry)

DE Human mismatch repair pathway protein, Mlh1.

KM Mismatch repair; MSH2; primer; identification; defect; alteration;

KW cancer; tumour; vaccine.

OS Homo sapiens.

PN WO9514085-A2.

PD 26-MAY-1995.

PF 17-NOV-1994; 94WO-US13385.

PR 13-JUN-1994; 94US-0259310.

PR 17-NOV-1993; 93US-0154792.

PR 07-DEC-1993; 93US-0163449.

XX (UYVE-) UNIV VERMONT & STATE AGRIC COLLEGE.
 PA (DAND) DANA FARBER CANCER INST.
 XX Fishel R, Kolodner RD, Reenan RAG;
 XX WPI: 1995-200377/26.
 DR N-PSDB: AAQ94015.
 XX
 PT Determining alteration in human mismatch repair pathways - used in
 PT the diagnosis, prognosis and therapy of cancers and in screening
 PT assays
 XX
 PS Example 13; Page 239-242; 256pp; English.
 CC
 CC AAR76071 is the human mismatch repair pathway protein Mlh1. Defects or
 CC alterations in a human mismatch repair gene results in the accumulation
 CC of unstable repeated DNA sequences, a feature of a number of different
 CC cancers. The identification of a defect in the mismatch repair pathway
 CC can be diagnostic of a predisposition to cancer and prognostic for a
 CC particular mammalian cancer e.g colorectal, ovarian, endometrial
 CC (uterine), renal, bladder, skin, rectal and bowel. The nucleotide
 CC sequences and polypeptides of the hMSH2 gene may also be used for
 CC therapy and in vaccines.
 CC
 SQ Sequence 756 AA:

Query Match 39.1%; Score 1452; DB 16; Length 756;
 Best Local Similarity 39.6%; Pred. No. 6, 5e-111;
 Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;

18 IRLESEVYVRIAGVGIORPSSAVKELIENSIDAGASSVAVKDGILQVSDGHC 77
 8 IRRLDTVYVRIAGVGIORPNAIKEMINCLDAKSTSIQYVEGGLIKLIQIDNGTG 67
 78 IRREDIALICERHTTSKLSAYEDLIOTIKSMGREGALSMYVGVVTVTTTIEGLHGR 137
 68 IRKEDIYICERPTTSKLOSFEDLASISYTRGRBALISISVAVHTTTTADGCKAYR 127
 138 VSTRGVYVMEKPEKPCAVKGTOWVENVFYNNVARKKTLQNSNDYPRIVDFISFAVHH 197
 128 ASYSOGKLAPKPCAGNCGQITVEDLFYNNATRRKALKMSEYKILEVGVSYVHN 187
 198 INVTESCKRHGANRADVSHASTSSRLDAIRSVYGVYVVDLIEIKVSTEDADADSTFKMDG 257
 188 AGISFSYVKKOGVETADVETLPMASTVDNIRISFIGNAVSREILEIGCEDKTA---FKMG 244
 258 YISNNAVYAKKITMLFENDRLNDCTALKAILEFYVSATLPOASKPPLYSIMHLPSEHVD 317
 245 YISNNAVYAKKITMLFENDRLNDCTALKAILEFYVSATLPOASKPPLYSIMHLPSEHVD 317
 318 VNIHPTKKEVSLNDEIRIETIRNAIEEKLMSNTTIFOTQALNTSGIAQANPQ--KDK 375
 305 VNVHPTKHEVHFLHESTLERVOQHIESKLSSNNMYFTQTL-LPGLAGPSGEVAKST 363
 376 VSEASMSGTSQKITPVOSWRTPDNPDSGLHTYMGSSNTL-----EKKFPLVS 426
 364 TSLTSSSTSSSDKYVAHOWRTRDSRQ--KLDAFLQPLSKPLSSQPAVTVTEKTDISS 421
 427 VRNVVR-----SRNKKDAGDIS-----444
 422 GRANQOQDEMLPAPAEVAAKNSLEGDTTKGTSEMSEKRGPTSSNPKRHREDSVEM 481
 445 -----SRHELLV-----EIDSPFGLDLYKVKCTVYGLADEAFALI 481
 482 VEDDSREKMAACPRRRRIINLTVSLQEIINBOGHEVIREMHNHSVGCVPOMALA 541
 482 OHMTRIYVNVVVISKEIMAQALCRGNFNAIQISEPAPDELLVMAKDEIMDEKO 541
 542 QHOTKLYLTLTTLSBELTYOIIYFANFGVRLSEPAFLPFLAMALADSESGTEED 601
 542 DEKLEIAEVTEILKENAEMINNEYFSIHIDOGSKLRLVAVDQTPMDRLPEVLAIG 601

DB 602 GPKREGIAEYIVFLKKAKEMLADYFSLIDEENGLGLEPLLDNTVPPLEGIPILRIA 661
 OY 602 NDYTMDEKECEFRVTASANGFYALHPILPNPSSGNGIHLKYKNDSDMADEHNDLSD 661
 DB 662 TEYNMDEKECEFRVTASANGFYALHPILPNPSSGNGIHLKYKNDSDMADEHNDLSD 661
 OY 662 ENDV---DQELIAEAAVAREMTLOHVLPSMRLFLKPKSNATDGTPOVASLEKLY 718
 DB 694 ESTLSGQSEVGSIPNSN---KWEIHYVYKALSHILPFRHFTEDENILQANLPDLY 750
 OY 719 KIFEGC 724
 DB 751 KVFERC 756

RESULT 7
 AAB85854
 ID AAB85854 standard; Protein: 756 AA.

AC AAB85854;

DR 29-OCT-2001 (first entry)

DE Human MLH1 protein.

KW Hypermutable bacteria; mismatch repair gene; MMR gene; Muth; Muts; Mult;
 KW Muty; PMS2; MLH1; MLH3; PMSR; biocatalysis; bioremediation; biochemical;
 KW drug discovery; detoxification; toxin; biotransformation.
 OS Homo sapiens.

PN W0200159092-A2.

PD 16-AUG-2001.

PF 12-FEB-2001; 2001MO-US04339.

PR 11-FEB-2000; 2000US-0181929.

PA (UYVO) UNIV JOHNS HOPKINS.

PI Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW;

DR WPI: 2001-514664/56.

DR N-PSDB: AAH76367.

XX Making hypermutable bacteria for biocatalysis, bioremediation and drug
 PT discovery, involves introducing polynucleotide comprising dominant
 PT negative allele of mismatch repair gene under regulatory sequence
 PT control

XX Example 1; Page 44; 68pp; English.

XX The invention provides a method for generating a hypermutable bacteria.
 CC The method involves introducing a polynucleotide having a dominant
 CC negative allele of a mismatch repair (MMR) gene under the control of an
 CC inducible transcription regulatory sequence, into a bacterium. The cell
 CC becomes inducibly hypermutable. The method is useful to create desirable
 CC output traits for commercial applications, using dominant negative
 CC alleles of mismatch repair proteins. The mismatch repair gene is a Muth,
 CC Muts, Mult or Muty homologue and can be selected from PMS2, MLH1, MLH3,
 CC PMSR or PMSR homologue. The hypermutable bacteria is useful for the
 CC production, biocatalysis, bioremediation and drug discovery. It is also
 CC useful in manufacturing industry for the generation of new biochemicals
 CC manufacturing processes or those used as catalysts, for remediation of
 CC toxins present in the environment including polychlorobenzene, heavy
 CC metals and other environmental hazards for which there is a need to
 CC remove them from the environment. The hypermutable bacteria is also
 CC useful for screening novel mutations in a gene or a set of genes that
 CC produce variant siblings that exhibit a new output trait not found in
 CC wild type cells. The bacteria are also useful for producing increased
 CC quantity or quality of protein or non-protein therapeutic molecule e.g.

CC Penicillin G, Erythromycin and Clavulanic acid, by biotransformation.
 CC Dominant negative alleles of the MMR gene are useful for producing higher
 CC quantities of recombinant polypeptides. The present sequence represents
 CC a human MLH1 protein.
 CC XX

SO Sequence 756 AA;

Query Match 39.1%; Score 1452; DB 22; Length 756;
 Best Local Similarity 39.6%; Pred. No. 6,5e-111;
 Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;

OY 18 IRLLEESVNRIRIAGVIRPSSAVKELIENSIDAGASSVAVKGGKLIQVSDGHC 77
 DB 8 IRLDETVNRIAGVIRPANAIREMENCIDAKSTSIQVYKGGKLIQIDQNGTG 67
 OY 78 IREDELAICERHTTSKLSAYEDQITKSGFGEALASMTYGVHTVTITTEGQLHGYR 137
 DB 68 IRKEDLDIVCERHTTSKLSQFEDLASISTYGFGEALASISHVAHTITTKTADGKAYR 127
 OY 138 VSTRDGMENEPKPCAAVKGTOYVENVLFYNNVARKKTLQNSNDYPKIYDFISRPVHH 197
 DB 128 ASYSDGKLKAPKPCAGNOGTQITVEDLFYNIATRRKALKNPSEYKILEVGYRSVHN 187
 OY 198 INTFECRKHGANKRADVHASTSSRIDAIRSVYGASVVDLITIKSYEDADASTIRKMG 257
 DB 188 AGISFSEVKKGQETVADYRFLPNASTVDNTRISIFGNVSRLEIGCEDKTLA---FKMG 244
 OY 258 YISNNAVYAKKTIIMLEIDRLVDCTALKRAIEFYVSATIPQAKPPIYKSHLPSEHYD 317
 DB 245 YISNNAVYAKKCFELFIFIRHIVESTSKRAITVYAAVLPKTHPFLYSLTSIQOND 304
 OY 318 VNIHPKKEVSLNDOERIIETRNALIEKIMNSNTTRIFQTOALNLSGLAAMPQ--KDK 375
 DB 305 VNVHPTKHEVHFHESILIEVQGHIESKLGSSNSRMFTOTL-LPGIAGSGEYVKST 363
 OY 376 VSSASGSGTKSKIPYVSOVWRTDPNPSGRHTTYHGGSSNL-----EKTEDLYS 426
 DB 364 TSTSSSTSGSSDKYVAHQWVRTDSRQ--KIDAFIQPLSKPLSOPQALVTFEDKTDISS 421
 OY 427 VRRNVV-----SRNQKADGLS----- 444
 DB 422 GRARODEEMLELPAPEVAAKNOSLEGDTTKTSEMSEKRGPTSSNPKRRHEDSDVEM 481
 OY 445 ---SHELLV-----EIDSFHPGLDITVKNCTYVGLADEAFALI 481
 DB 482 VEDDSRKEMTACTPRRIINILTSVLSLQEIENEQGHVLRKMLHNSFEVGVNPOWALA 541
 OY 482 QHNTRIYVNVVNIKELWQALCRFGNNAIQISEPARLQELLMALKDDDELMSDEKD 541
 DB 542 QHOTKLYLNTTKLSEELFYQILIDPANGVLRSEPAFLDLAMALADSPSGWTEED 601
 OY 542 DEKLTAIVNTLTKENAEINNEYFISIHIDODGKLTLPVLPVLDQYTPMDRLDEFFVALG 601
 DB 602 GREGLAETIVELFKKAEMADYFSELEIDEEGNLIGPLIDINNYPPLEGILFIIRLA 661
 OY 602 NDVTWDEKECFETVAVSAVGNFYALHPILPNPSGNGIHLYKKNRDSMADEHAENLISD 661
 DB 662 TEVNMDEKECFESLSKEAMFYSI-----RKQYISE 693
 OY 662 ENDV---DDELLAEAPAAQOREMTIQHVLFPSEMLFLKPKKMAIDGTFVQVASELKY 718
 DB 694 ESTLSGOOSVPGSIPNSW---KMTVEHIVYKALRSHIILPKHFTEDGNITLQANLPDLX 750
 OY 719 KIPERC 724
 DB 751 KYFERC 756

RESULT 8
 AAG63956
 ID AAG63956 standard; protein; 756 AA.
 XX
 AC AAG63956;

XX 29-OCT-2001 (first entry)
 XX Amino acid sequence of human mismatch repair protein MLH1.
 DE
 DE MLH1, mismatch repair gene; MMR gene; hypermutable yeast.
 KW
 OS Homo sapiens.
 OS
 XX W0200162945-A1.
 XX
 PD 30-AUG-2001.
 XX
 XX 21-FEB-2001; 2001MO-US05447.
 XX
 XX 23-FEB-2000; 2000US-0184336.
 XX
 PA (UYJO) UNIV JOHNS HOPKINS.
 PA (NICO/) NICOLAIDES N C.
 PA (SASS/) SASS P M.
 PA (GRAS/) GRASSO L.
 PA (VOGE/) VOGELSTEIN B.
 PA (KINZ/) KINZLER K W.
 XX
 PI Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW;
 XX
 DR WPI; 2001-522820/57.
 DR N-PSDB; AAH75044.

PT Making hypermutable yeast that exhibit novel selected output traits for
 PT commercial applications, comprises introducing polynucleotide
 PT containing dominant negative allele of mismatch repair gene -
 PS Disclosure; Page 43-44; 60pp; English...

CC The present sequence represents human MLH1. MLH1 is a mismatch repair
 CC (MMR) gene. The specification describes a method for making a
 CC hypermutable yeast, comprising introducing a polynucleotide containing
 CC a dominant negative allele of a mismatch repair (MMR) gene, into a
 CC yeast, whereby the cell becomes hypermutable. The method is useful
 CC to create desirable output traits for commercial applications, using
 CC dominant negative alleles of mismatch repair proteins. The hypermutable
 CC yeast is useful for production, biocatalysis, bioremediation and drug
 CC discovery. It is also useful in genetic screens for the direct selection
 CC of variant subclones that exhibit new output traits. The hypermutable
 CC yeast is also useful in the manufacturing industry for the generation
 CC of new biochemicals, for detoxifying noxious chemicals from by-products
 CC of manufacturing processes or those used as catalysts, for remediation
 CC of toxins present in the environment including polychlorobenzenes, heavy
 CC metals and other environmental hazards for which there is a need to
 CC remove them from the environment. The yeast is also useful for producing
 CC increased quantity or quality of protein or non-protein therapeutic
 CC molecule e.g., Penicillin G, Erythromycin and Clavulanic acid, by
 CC biotransformation.

SO Sequence 756 AA;

Query Match 39.1%; Score 1452; DB 22; Length 756;
 Best Local Similarity 39.6%; Pred. No. 6,5e-111;
 Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;

OY 18 IRLLEESVNRIRIAGVIRPSSAVKELIENSIDAGASSVAVKGGKLIQVSDGHC 77
 DB 8 IRLDETVNRIAGVIRPANAIREMENCIDAKSTSIQVYKGGKLIQIDQNGTG 67
 OY 78 IREDELAICERHTTSKLSAYEDQITKSGFGEALASMTYGVHTVTITTEGQLHGYR 137
 DB 68 IRKEDLDIVCERHTTSKLSQFEDLASISTYGFGEALASISHVAHTITTKTADGKAYR 127
 OY 138 VSTRDGMENEPKPCAAVKGTOYVENVLFYNNVARKKTLQNSNDYPKIYDFISRPVHH 197
 DB 128 ASYSDGKLKAPKPCAGNOGTQITVEDLFYNIATRRKALKNPSEYKILEVGYRSVHN 187

```

OY 198 INVTSCRRKGANRADVHASTSSRLDAIRSVYGASVYVDLEIKVSEDAADSIKMDG 257
Db 198 AGISFSVYKOGETVADVPTLNPASTVDNIRISIFGNVARELIEGCEKTLA---FKMG 244
OY 258 YISNANYVAKKITMLIFINDRLVCTALKAIEFYVSTLTQASKPTIYMSIHLPSEHYD 317
Db 245 YISNANYVAKKITMLIFINDRLVCTALKAIEFYVSTLTQASKPTIYMSIHLPSEHYD 317
OY 318 VNIHPTKEVSLINOERLIETIRNAIEKILMSNTTRIFOTQALNLSGIAQANQ--KDK 375
Db 305 VNIHPTKEVSLINOERLIETIRNAIEKILMSNTTRIFOTQALNLSGIAQANQ--KDK 375
OY 376 VSEASMSGTSOKIPIVSQAVRTDPNPSGRRLTHYHGQSSNL-----EKKFDLVS 426
Db 364 TSLTSSSTGSSSKVYAHQVWRTDSRQ--KLDAFLQPLSKPLSSQPAIVTEDKIDISS 421
OY 427 VNNVVR-----SRNOKDAGDLS-----444
Db 422 GRANOODEMELPAPAEVAAKNOSLEGDTTKGTSENSEKRGPTSSNPKRRHREDSDVEM 481
OY 445 -----SRHELLY-----EIDSFHPGLDLYKNTCTYVGLADEAFALI 481
Db 482 VEDDSRREKMTACTPRRRIINTLSVLSLOEINEGHEVLREMLNHNHSFVGCVPQWALA 541
OY 482 QHNTRLVYVNVNISKELMYOALCRFGNNAIOLSEPAIOLLYMALKDELSDEKD 541
Db 542 QHOTKLYLNTKLSBELFYOLITDFANGVRLSEPAIOLLYMALADSPESGWTED 601
OY 542 DEKLEIAEVNTEILKENAEMINEFSIHIDQKLTRLPVVLDQVPPMDRLPEFVALG 601
Db 602 GPRGELIAYIEVLEIKKAEMLADYFSLIDEBSNLGLPLLDNVPPLIEGLPIFILRLA 661
OY 602 NDVTWDEKCEPRTVAVSAVGNFYALHPILPNPSGNGIHLTKNRDSMADEHAENDLSD 661
Db 662 TEVNMDEKCEPRTVAVSAVGNFYALHPILPNPSGNGIHLTKNRDSMADEHAENDLSD 661
OY 662 ENDY---DOELLAEAEMAAOEWITQHVLPFSMRLFLKPPKSMATDGFYOVASLEKIX 718
Db 694 ESTLSQOGEVEGSIPIPSN---KWVEHIVYKALSHILPHEFDGNIJLOLANLPDLY 750
OY 719 KIFERC 724
Db 751 KYFERC 756

```

```

Db 188 AGISFSVYKOGETVADVPTLNPASTVDNIRISIFGNVARELIEGCEKTLA---FKMG 244
OY 258 YISNANYVAKKITMLIFINDRLVCTALKAIEFYVSTLTQASKPTIYMSIHLPSEHYD 317
Db 245 YISNANYVAKKITMLIFINDRLVCTALKAIEFYVSTLTQASKPTIYMSIHLPSEHYD 317
OY 318 VNIHPTKEVSLINOERLIETIRNAIEKILMSNTTRIFOTQALNLSGIAQANQ--KDK 375
Db 305 VNIHPTKEVSLINOERLIETIRNAIEKILMSNTTRIFOTQALNLSGIAQANQ--KDK 375
OY 376 VSEASMSGTSOKIPIVSQAVRTDPNPSGRRLTHYHGQSSNL-----EKKFDLVS 426
Db 364 TSLTSSSTGSSSKVYAHQVWRTDSRQ--KLDAFLQPLSKPLSSQPAIVTEDKIDISS 421
OY 427 VNNVVR-----SRNOKDAGDLS-----444
Db 422 GRANOODEMELPAPAEVAAKNOSLEGDTTKGTSENSEKRGPTSSNPKRRHREDSDVEM 481
OY 445 -----SRHELLY-----EIDSFHPGLDLYKNTCTYVGLADEAFALI 481
Db 482 VEDDSRREKMTACTPRRRIINTLSVLSLOEINEGHEVLREMLNHNHSFVGCVPQWALA 541
OY 482 QHNTRLVYVNVNISKELMYOALCRFGNNAIOLSEPAIOLLYMALKDELSDEKD 541
Db 542 QHOTKLYLNTKLSBELFYOLITDFANGVRLSEPAIOLLYMALADSPESGWTED 601
OY 542 DEKLEIAEVNTEILKENAEMINEFSIHIDQKLTRLPVVLDQVPPMDRLPEFVALG 601
Db 602 GPRGELIAYIEVLEIKKAEMLADYFSLIDEBSNLGLPLLDNVPPLIEGLPIFILRLA 661
OY 602 NDVTWDEKCEPRTVAVSAVGNFYALHPILPNPSGNGIHLTKNRDSMADEHAENDLSD 661

```

N-PSDB: AAD39200.

Making hypermutable antibody-producing cells for producing antibodies with e.g. enhanced biochemical activity, comprises introducing into a cell a polynucleotide with a dominant negative allele of a mismatch repair gene.

Disclosure: Page 18; 75pp; English.

The invention relates to a method for making a hypermutable, antibody-producing cell. The method comprises introducing into a cell, which is capable of producing antibodies, a polynucleotide comprising a dominant negative allele of a mismatch repair gene (MMR). The method is useful for generating genetically altered antibody-producing cell lines with improved antibody characteristics. In particular, the method is useful for generating genetic diversity within immunoglobulin genes directed against an antigen to produce antibodies with enhanced biochemical activity or for generating antibody-producing cells with increased level of antibody production. The present sequence is human mismatch repair protein, MLH1.

Sequence 756 AA;

Query Match 39.1%; Score 1452; DB 23; Length 756; Best Local Similarity 39.6%; Pred. No. 6.5e-111; Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;

18 IRRLSESVNRIAGVYVORPSSAVKELIENSIDGASVYAVKDGKLTIOYSDGHC 77

8 IRRIDEVYVNRHIAAGEVYVORPANAIEKMEIENCDAKSTIOVYKEGGLTIOQDNGT 67

78 IRFDLILICRHTTSLSAVEDLTQTKSGFGEALASMTYGVHVTITTEGOLHGR 137

68 IRKEDLVICERFTSKLQSFEDLASISTYGFREALASISHVAHVTITTKADGCAVR 127

138 VSYRQVMEVNEPKCAVKTQVAVENLFYVNAVARKKTIONSDQVPIYDFISRAVH 197

128 ASYSDGKLAPKPCACANQGTQITVEDLFYNIAIRKALKNPSEEGKILLEVGRSVN 187

198 INVTFSCRRKGANRADVHASTSSRLDAIRSVYGASVYVDLEIKVSEDAADSIKMDG 257

188 AGISFSVYKOGETVADVPTLNPASTVDNIRISIFGNVARELIEGCEKTLA---FKMG 244

258 YISNANYVAKKITMLIFINDRLVCTALKAIEFYVSTLTQASKPTIYMSIHLPSEHYD 317

245 YISNANYVAKKITMLIFINDRLVCTALKAIEFYVSTLTQASKPTIYMSIHLPSEHYD 317

318 VNIHPTKEVSLINOERLIETIRNAIEKILMSNTTRIFOTQALNLSGIAQANQ--KDK 375

305 VNIHPTKEVSLINOERLIETIRNAIEKILMSNTTRIFOTQALNLSGIAQANQ--KDK 375

376 VSEASMSGTSOKIPIVSQAVRTDPNPSGRRLTHYHGQSSNL-----EKKFDLVS 426

364 TSLTSSSTGSSSKVYAHQVWRTDSRQ--KLDAFLQPLSKPLSSQPAIVTEDKIDISS 421

427 VNNVVR-----SRNOKDAGDLS-----444

422 GRANOODEMELPAPAEVAAKNOSLEGDTTKGTSENSEKRGPTSSNPKRRHREDSDVEM 481

445 -----SRHELLY-----EIDSFHPGLDLYKNTCTYVGLADEAFALI 481

482 VEDDSRREKMTACTPRRRIINTLSVLSLOEINEGHEVLREMLNHNHSFVGCVPQWALA 541

482 QHNTRLVYVNVNISKELMYOALCRFGNNAIOLSEPAIOLLYMALKDELSDEKD 541

542 QHOTKLYLNTKLSBELFYOLITDFANGVRLSEPAIOLLYMALADSPESGWTED 601

542 DEKLEIAEVNTEILKENAEMINEFSIHIDQKLTRLPVVLDQVPPMDRLPEFVALG 601

602 GPRGELIAYIEVLEIKKAEMLADYFSLIDEBSNLGLPLLDNVPPLIEGLPIFILRLA 661

602 NDVTWDEKCEPRTVAVSAVGNFYALHPILPNPSGNGIHLTKNRDSMADEHAENDLSD 661

DB 662 TEVNMEKECFESLSEKAMFYSI-----RKQYISE 693
 QY 662 ENDV---DDELLAEAAEAQAQREMTIOHVLFPMSRLFLAPPKSMATDGTFFVQVASEKLY 718
 DB 694 ESTLSGQSEVPSPINPSW---KWTVEHIVYKALRSHIILPPKHFTEBDGNILQIANLPDLY 750
 QY 719 KIFERC 724
 DB 751 KIFERC 756

RESULT 10
 AAU98778
 ID AAU98778 standard; Protein: 756 AA.
 AC AAU98778;
 XX
 XX 24-SEP-2002 (first entry)
 DT
 XX Human Mult. homologue, MLH1.
 DE
 XX Human; DNA mismatch repair enzyme; MLH1;
 KW Mut. L. homologue; dominant negative mutation; antibacterial;
 KW hypermutable cell; post meiotic segregation increased; PMS1; PMS2;
 KW Mult. homologue 2; MSH2; MLH2; antimicrobial.
 XX Homo sapiens.
 OS
 XX WO200238750-A1.
 PN
 XX 16-MAY-2002.
 PD
 XX 07-NOV-2000; 2000WO-US30587.
 PF
 XX 07-NOV-2000; 2000WO-US30587.
 PR
 XX 07-NOV-2000; 2000WO-US30587.
 PA (MORP-) MORPHOTEK INC.
 XX
 XX Grasso L, Nicolaides NC, Sass PM;
 PI
 XX WPI: 2002-508210/54.
 DR N-PSDB; ABR66091.
 DR
 XX Making mammalian cell hypermutable for obtaining a mammalian cell that
 PT is resistant to selected microbe by introducing polynucleotide
 PT comprising dominant-negative allele of mismatch repair gene into
 PT mammalian cell
 PS
 PS Disclosure; Page 18; 68pp; English.

CC nucleotide sequence of the gene or mRNA transcribed from the gene, a
 CC protein encoded by gene or its phenotype. The composition
 CC is useful for identifying antimicrobial agents, microbe-specific toxic
 CC molecules, and for producing new phenotypes of the cell. (M1) is useful
 CC for creating genetically altered antimicrobial molecules, and also for
 CC creating cell lines that manufacture antimicrobial molecules for use in
 CC large scale production of antimicrobial agents for clinical studies.
 CC (M1) is also useful in cell lines that express known antimicrobial
 CC agents to enhance the biochemical activity of the antimicrobial agent.
 CC The present sequence represents human MLH1.
 CC
 XX
 XX Sequence 756 AA:
 SO
 Query Match 39.1%; Score 1452; DB 23; Length 756;
 Best Local Similarity 39.6%; Pred. No. 6.5e-111;
 Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;

QY 18 IRLLESYVNRINAAAGEYIOPSSAVKELIENSIDAGASSVAVKDGGLIYVSDDGHG 77
 DB 8 IRLDETVNRINAAAGEYIOPSSAVKELIENSIDAGASSVAVKDGGLIYVSDDGHG 77
 QY 78 IRLDETVNRINAAAGEYIOPSSAVKELIENSIDAGASSVAVKDGGLIYVSDDGHG 77
 DB 68 IRLDETVNRINAAAGEYIOPSSAVKELIENSIDAGASSVAVKDGGLIYVSDDGHG 77
 QY 138 VSYRDGEMENEPKCAVKTQVAVENLFYNNVARKKTLONSDDPKIYDTSRPAVH 197
 DB 128 ASYSDGKLKAPKPCAGNQTQITVEDLFYNNVARKKTLONSDDPKIYDTSRPAVH 197
 QY 198 INTFSCRKHGRADYHSAHSTSRDLAISVYGVADYDLIELKYSIEDAADSIFMDG 257
 DB 188 AGISFVKKQGTADVADRLPNASVYDNIKISIFGNAVSRELIEGCEKDTLA--FPMG 244
 QY 258 YISNANYVAKKIMILEFINRLVDCALRAIEFVSATLPQASKPPIYVSIHLPSEVD 317
 DB 245 YISNANYVAKKIMILEFINRLVDCALRAIEFVSATLPQASKPPIYVSIHLPSEVD 317
 QY 318 VNIHPTKREYSLNQEIIETIRNAIEEKLNSNTTRIFQTOALNLSGIAQNPQ--KDK 375
 DB 305 VNIHPTKREYSLNQEIIETIRNAIEEKLNSNTTRIFQTOALNLSGIAQNPQ--KDK 375
 QY 376 VSEASMGSGTKSOKIPYSONVTRDPRNSGRLLTHYHGGSSNL-----EKEDLV 426
 DB 364 TSLTSSSTSGSDKYVAHQVTRDPRNSGRLLTHYHGGSSNL-----EKEDLV 426
 QY 427 VSNVVR-----SRNOKDAGDLS----- 444
 DB 422 GRARQDEEMLELPAPAEVAKNOSLEGDTKGTSESEKRGPTSSNRRHREDSDEM 481
 QY 445 -----SRHELY-----EIDSFHFGLLDIYKNCYVGLADEAFALI 481
 DB 482 VEDDSKREMAACTPRRIINISVLSLOEINOGHEVIREMHNHSPGCNPNPQALA 541
 QY 482 QHNTRILYVNVNISKELMTOALCRGNFNAIOLSEPADLQELLYVALKDELMSKDK 541
 DB 542 QHNTRILYVNVNISKELMTOALCRGNFNAIOLSEPADLQELLYVALKDELMSKDK 541
 QY 542 QHNTRILYVNVNISKELMTOALCRGNFNAIOLSEPADLQELLYVALKDELMSKDK 541
 DB 542 QHNTRILYVNVNISKELMTOALCRGNFNAIOLSEPADLQELLYVALKDELMSKDK 541
 QY 542 DEKLEIAEVNTELLKEAENINEFSTHIDODKTRFLPVILQYVPMRRLPEFVALG 601
 DB 602 GPEGLAEIVTEVFLKKAEMADYFSLDEBGLGLPLIDYVYPLGLPFIIRLA 661
 QY 602 NDYTWDEKECFEFTVASAVNGFYALHPILPNPSGNGIHLKKNRSMADHEANLISD 661
 DB 662 TEVNMEKECFESLSEKAMFYSI-----RKQYISE 693
 QY 662 ENDV---DDELLAEAAEAQAQREMTIOHVLFPMSRLFLAPPKSMATDGTFFVQVASEKLY 718
 DB 694 ESTLSGQSEVPSPINPSW---KWTVEHIVYKALRSHIILPPKHFTEBDGNILQIANLPDLY 750
 QY 719 KIFERC 724
 DB 751 KIFERC 756

RESULT 11
ID AAG75536 standard; Protein: 775 AA.
AC AAG75536;
DT 03-SEP-2001 (first entry)
DE Human colon cancer antigen protein seq ID NO:6300.
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 3.
OS Homo sapiens.
PN W0200122920-A2.
PD 05-APR-2001.
PF 28-SEP-2000; 2000MO-US26524.
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI: 2001-235357/24.
XX N-PSDB; AAH34941.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 11; Page 7756-7760; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens (P), where
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing P.
XX Additionally, N may be used to produce the colon cancer-associated P,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX present invention.
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 775 AA:
Query Match 39.18; Score 1452; DB 22; Length 775;
Best Local Similarity 39.66; Pred. No. 6,8e-111;
Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;
OY 18 IRRLEESVNRIRAGEYIORSSAVKELIENSIDAGASSVAVKDGKLIQVSDGNG 77
DB 27 IRRLEEVNRIRAGEYIORSSAVKELIENSIDAGASSVAVKDGKLIQVSDGNG 77
OY 78 IRRLEEVNRIRAGEYIORSSAVKELIENSIDAGASSVAVKDGKLIQVSDGNG 86
DB 87 IRRLEEVNRIRAGEYIORSSAVKELIENSIDAGASSVAVKDGKLIQVSDGNG 137
OY 138 VSRDGVNEMEPKCAAVGTQVMEVNTLYNMVARKKTLQNSDDYDKLVDISFPAVH 197
DB 138 VSRDGVNEMEPKCAAVGTQVMEVNTLYNMVARKKTLQNSDDYDKLVDISFPAVH 197
PA (HUMA-) HUMAN GENOME SCI INC.

147 ASYSDCKLAPKPPKCAQNOCTQITVEDLFFYNATRRKALKNPSEEGKILLEVGRYSVN 206
OY 198 INTFSCRKRGANRADVSHASTSSRLDAIRSYVAGSVVDLEIKYVEDADASTFKDG 257
DB 207 AGISFSYVKQGEVADVRLPNASTVDNRSIFGNVNSHELIECEDKTLA--FKNG 263
OY 258 YISNANYVAKRTIMLEIFINDRLVDCAIKRAIEFYASATLPQASKPFTYMSIHJSEHVD 317
DB 264 YISNANYVAKRTIMLEIFINDRLVDCAIKRAIEFYASATLPQASKPFTYMSIHJSEHVD 317
OY 318 YNHHPKKEVSLINDERTIETIRNAIEERKLNMSNTTRIFQALNLSGIAQANQ--KDK 375
DB 324 VNVHPKHEVHLHEESLIERVOQHIESLIGNSNRNYFQTL-LGLAIPSEMYKST 382
OY 376 VSEASMSGSGTKSOKIPIVSONVTPDRPNRSGRLHYTHGSSNL-----EKKFDLVS 426
DB 383 TSLTSSSTSGSDKVAHOMVKTDSRQ--KLDAFLPLSKPLSSQPAIVTEDEKTISS 440
OY 427 VNVNVR-----SRNOKDAGLS-----444
DB 441 GRARQODEMLELPAPAEVAAKNOSLEBDITKGTSESEKRGPTSSNPRKRRHEDSDVEM 500
OY 445 -----SRHELY-----EIDSSFRGLDIYKNTYGLADEAPALI 481
DB 501 VEDSRKEMTACPPRRRIINLTVSLLOEIRNBOGHEVLRMLNHSFVGCYNPQWALA 560
OY 482 QHNRLVLYNVVNSKELMVOALCREGNFNAIOLSEPAFLQELVWALKDELMSDEKD 541
DB 561 OHOTKVLNLTAKSEELFYOLITDPANGVRLSEPAFLQELVWALKDELMSDEKD 541
OY 542 DEKLEIAVNTLEIKENAMINEFSLIHDOGDKLRLPVLDQVTPMDRLPEFLAG 601
DB 621 GPKGLAEVYELFKKRAEMADYFSLIDEENGLGLFLIDNVYPRLEGPIFILRA 680
OY 602 NDVTWDEKCEFTVVASAVGFYALHPILPNPSGNGHLLKKNDSMADEHENDLSD 661
DB 681 TEVWVDEKCEFTVVASAVGFYALHPILPNPSGNGHLLKKNDSMADEHENDLSD 661
OY 662 ENDV-----DOELLAFAEANAQREWTIOHVLFPSSMLFLKPPKSMATDGTVOVASLEKTY 718
DB 713 ESTLSQOOSVPSIPSNW---KWVEHIVYALNSHILPPHFTEDGNITQLANLPDLY 769
OY 719 KIFERC 724
DB 770 KYFERC 775
RESULT 12
ID AAR79008 standard; Protein: 756 AA.
AC AAR79008;
DT 26-MAR-1996 (first entry)
DE Human DNA repair protein hMLH1.
KW DNA repair protein; hMLH1; hMLH2; hMLH3; therapy; cancer; vectors;
KW DNA synthesis; diagnosis; disease; mutL4.
OS Homo sapiens.
PN W09520678-A1.
PD 03-AUG-1995.
PE 25-JAN-1995; 95MO-US01035.
PR 23-AUG-1994; 94US-0294312.
PR 27-JAN-1994; 94US-0187757.
PR 16-MAR-1994; 94US-0210143.
PA (HUMA-) HUMAN GENOME SCI INC.

XX Adams MD, Fleischmann RD, Fraser CM, Fuldner RA.
PI Haseltine WA, Kirkness EF, Rosen CA, Ruben SM, Wei Y;
XX WPI: 1995-275461/36.
DR N-PSDB: AAQ97525.
XX
PT Polynucleotide(s) encoding human multi homologues, hMLH1, hMLH2 and
XX hMLH3 - used for therapeutic treatment of, e.g. hereditary cancer
XX
PS Claim 8; Figure 1; 124pp; English.
XX
CC The polynucleotides described in AAQ97525-27 encode the human
CC analogues of the prokaryotic multi DNA repair gene. The polypeptides
CC they encode (AAQ97508-R/9010) are used for therapeutic purposes e.g.
CC in the treatment of cancer, esp. hereditary cancer. They may also
CC be used for in vitro manipulation of DNA, synthesis of DNA and the
CC manufacture of DNA vectors and in methods of diagnosing a disease or
CC a susceptibility to a disease related to a mutation in the hMLH1, -2
CC or -3 DNA repair genes.
XX
SQ Sequence 756 AA:

Query Match 38.8%; Score 1439; DB 16; Length 756;
Best Local Similarity 39.4%; Pred. No. 7.8e-110;
Matches 110; Conservative 144; Mismatches 216; Indels 116; Gaps 11;

18 IRLLEESVNRILAGEVIRPSSAVKELIENSIDAGASSVAVKDGKLIQVSDG 77
18 IRLLEESVNRILAGEVIRPSSAVKELIENSIDAGASSVAVKDGKLIQVSDG 77
8 IRLDETVNRILAGEVIRPSSAVKELIENSIDAGASSVAVKDGKLIQVSDG 67
78 IRLDETVNRILAGEVIRPSSAVKELIENSIDAGASSVAVKDGKLIQVSDG 137
68 IRLDETVNRILAGEVIRPSSAVKELIENSIDAGASSVAVKDGKLIQVSDG 127
138 VSYRQGVMEPEPCAAVGTGVTGVTGVTGVTGVTGVTGVTGVTGVTGVTG 197
128 ASYSDGKLAAPKPCAGNOGTOTLTEDLTNTATRRALKNPSEYKLTLEVGRYSV 187
198 INVTFSCRRKGANRADVHASTSSRLDAIRSVYGVASVVDLEIKVSEDAADSIK 257
188 AGISFVKKOGVADVRILPNSATVDNIRSVGNVSRLEIGEDKTLA---EKMG 244
258 YISNANYAKKTMILFINDRLVDTALKRAIEFVSATLPQASKPTIYWSIHLPSE 317
245 YISNANYAKKTMILFINDRLVDTALKRAIEFVSATLPQASKPTIYWSIHLPSE 304
318 VNIHPKKEVSLINORIEITRNALTEKLMNSNTIRIOTOLANLSGIAQANPQ--K 375
305 VNIHPKKEVSLINORIEITRNALTEKLMNSNTIRIOTOLANLSGIAQANPQ--K 363
376 VSEASMSGSTKSQKIPVSNVTRTPRNPGRRLHTVWHGSSNT-----EKKF 426
364 TSLTSSSTGSSSDKYAHQWNRDSDRQ--KIDAFIQLPLSKPLSOPQALVTEDK 421
427 VNRNVR-----SRNQDAGDLS-----444
422 GRARQDDEMLLPAPAEVAAKQSLGDTTKGTSMESEKRCPTSSNPRKRHRSD 481
445 -----SRNELLV-----EIDSSPHGLDLYVKNCTYVGLADEAFALI 481
482 VEDDSKKEVTAACTPRRIINLTSLVSLQELNEGHEVLRNHNHSPGCVNPMWALA 541
482 OHNTRLYLNVNVIKELMQOALCRFQNFNAIOLSEAPLQELVYALKDDLEMSDEK 541
542 QHOTKYLINTKLSLELYQILYDFANFVRLSEPARLPFLMALALDPSGTEED 601
542 DEKLELAENVTEILKENAMENEFYSHIDQSKLRLPVLVDQYTPDMDRLPEFVAL 601
602 GPREGIAEYVEFLKKAEMADYFSLIEDBGNLIGLPLIDNVYVPLEGLPFIILRLA 661
602 NDVTWDEKCEPRTVASVNGNFYALHPPLIPNSGNGIHLIKNNRSMADENHENDLIS 661

DB 662 TEVNWDEKCEPRTVASVNGNFYALHPPLIPNSGNGIHLIKNNRSMADENHENDLIS 693
QY 662 ENDV---DDELTAFAEAMARQEMTIOHVLFPSPMRFLKPKSMATDGTFOVASLEKLY 718
DB 694 ESTLSGQSQSEVPISIRMSW---KMTVEHIYKALRSIILPKHFTEDGNILQLANLPDL 750
QY 719 KIFERC 724
DB 751 KIFERC 756

RESULT 13

ABB65767
ID ABB65767 standard; Protein; 664 AA.

AC ABB65767;
XX 26-MAR-2002 (first entry)

DT Drosophila melanogaster polypeptide SEQ ID NO 24093.
XX

DE Drosophila melanogaster polypeptide SEQ ID NO 24093.
XX

KW Drosophila: developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX

OS Drosophila melanogaster.
XX

PN WO200171042-A2.
XX

PD 27-SEP-2001.
XX

PF 23-MAR-2001; 2001MO-US09231.
XX

PR 23-MAR-2000; 2000US-191637P.
XX

PR 11-JUL-2000; 2000US-0614150.
XX

PA (PEKE) PE CORP NY.
XX

PI Venter JC, Adams M, Li PMD, Myers EW;
XX N-PSDB: ABL09870.
XX

DR WPI: 2001-656860/75.
XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -

PS Disclosure; SEQ ID NO 24093; 21pp + Sequence Listing; English.
XX

CC The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX sequences (AB17737-AB17072).
XX

CC The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 664 AA:
XX

Query Match 34.0%; Score 1261; DB 22; Length 664;
Best Local Similarity 36.9%; Pred. No. 3.3e-95;
Matches 270; Conservative 145; Mismatches 223; Indels 94; Gaps 11;

14 EPPRIRLEESVNRILAGEVIRPSSAVKELIENSIDAGASSVAVKDGKLIQVSD 73
6 OPVIRIRLEESVNRILAGEVIRPSSAVKELIENSIDAGASSVAVKDGKLIQVSD 65

74 DGGIRLEESVNRILAGEVIRPSSAVKELIENSIDAGASSVAVKDGKLIQVSD 133
66 NGTIRLEESVNRILAGEVIRPSSAVKELIENSIDAGASSVAVKDGKLIQVSD 125

QY 134 HGIVSYRDGVMEVNEPPCAAVKGTQVWVNFYNNVARKKTLQNSNDYPRIVETISRF 193
 DB 126 CGYKATYADGKIQGPPRCAGNOSTITCIDLYNNMFORQALRSAPAEFORISVIATAY 185
 QY 194 AVHINWTFESCRKIGANRADVHASTSSRLDAIRSYGASVVDLEIKYSEDADNDFE 253
 DB 186 AVHNRFVGTFLRKQDQAPALRTFVASSRSSENIITIGAAISKELEF-----SHDEYV 240
 QY 254 KMDG--YISNANYAKTITMLIFINDRLVDCALKRAIEFYVSAITLPOASKPEIYSIHL 311
 DB 241 KEFAECILTOVYSKCKOMLFTINORIVESTALRTSVDSIYATYLRGHHPEYMSLTL 300
 QY 312 PSHVDVNIHPKKEVSLNOERIEETIRNAIEEKLMSNTTRIFQOALNSGIAQANP 371
 DB 301 PPONDLVNHPKHEVHLYGEEIVDSIKOOVEANLLGSNATRFYFQ--LRLPA-----P 355
 QY 372 QKDVSASMGSGTSQKIPVSAWVTRDP-----RNPGRILTYWHGQSSNL 418
 DB 356 DLEETOLAD-----KTQRTYKEMVNTDSTDEQDKFLAPLYKSDSGVSSSSQSEASRLP 410
 QY 419 EKKFDLVSVNVRSSRNQKADGLSSRHLLIVEIDSSFPGLDIYKNCYVGLADEAF 478
 DB 411 EESFRVIAK--KSREVR-----LSSVDMKRREVEROCQVOLRSTLNLVYGCVDERR 462
 QY 479 ALIQHNTRLVNVNISKELMVOALCFRGNFNAITLSEPAILOELLYMALKDELMSD 538
 DB 463 ALFOHETRLVMCNRFSSEELFYORMYERFONCEITISPLPKRLILISEBAAGWT 522
 QY 539 EKDEKELIAVNTTEILKENAMENEFYSHIDQKTLRPVVDQYPRDMLPEVYL 598
 DB 533 PEDGKALADGADILLIKKAPIRVEYFGLRISDGLLESLPSELLHQHRCVAHLPYLL 582
 QY 599 ALGNDVYTWDEKECPRYVASVNGFYALHPPLIPNSGNGIHLYKKNRDSMADEHAENDL 658
 DB 583 RLATFVDMEOETRCFETFCRETARFY----- 608
 QY 659 ISDENVDVQELIAEAMAAO-----REWTOHYVLEPSPMRFLPKPKSMATDGTFFOVA 712
 DB 609 -----AQIDMREGATPAGFGRWTHMEVLEPPAFKYLPPRIKDO--IYEIL 652
 QY 713 SLEKTYIERC 724
 DB 653 NLEFLYKVERC 664
 RESULT 14
 AAB85849 standard; protein: 769 AA.
 ID AAB85849;
 AC AAB85849;
 XX 29-OCT-2001 (first entry)
 DT 29-OCT-2001 (first entry)
 DE Yeast MHL1 protein.
 XX Hypermutable bacteria; mismatch repair gene; MMR gene; MutH; MutS; MutL;
 KW MutY; PMS2; MHL1; MHL3; PMSR; biocatalysis; bioremediation; biochemical;
 RV drug discovery; detoxification; toxin; biotransformation.
 XX Saccharomycetes cerevisiae.
 OS
 PN WO200159092-A2.
 XX 16-AUG-2001.
 PD 12-FEB-2001; 2001WO-US04339.
 PF 11-FEB-2000; 2000US-0181929.
 XX (UYJO) UNIV JOHNS HOPKINS.
 PA Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW.
 PI

XX WP1: 2001-514664/56.
 DR N-ESDB; AAH76362.
 XX Making hypermutable bacteria for biocatalysis, bioremediation and drug
 PT discovery, involves introducing polynucleotide comprising dominant
 PT negative allele of mismatch repair gene under regulatory sequence
 PT control
 XX Example 1; Page 37-38; 68pp; English.
 PS The invention provides a method for generating a hypermutable bacteria.
 XX The method involves introducing a polynucleotide having a dominant
 CC negative allele of a mismatch repair (MMR) gene under the control of an
 CC inducible transcription regulatory sequence, into a bacterium. The cell
 CC becomes inducibly hypermutable. The method is useful to create desirable
 CC output traits for commercial applications, using dominant negative
 CC alleles of mismatch repair proteins. The mismatch repair gene is a MutH,
 CC MutS, MutL or MutM homologue. The hypermutable bacteria is useful for the
 CC PMSR or PMSR homologue. The hypermutable bacteria is useful for the
 CC production, biocatalysis, bioremediation and drug discovery. It is also
 CC useful in manufacturing industry for the generation of new biochemicals
 CC useful for detoxifying noxious chemicals from by-products of
 CC manufacturing processes or those used as catalysts, for remediation of
 CC toxins present in the environment including polychlorobenzene, heavy
 CC metals and other environmental hazards for which there is a need to
 CC remove them from the environment. The hypermutable bacteria is also
 CC useful for screening novel mutations in a gene or a set of genes that
 CC produce variant siblings novel mutations in a new output trait not found in
 CC wild type cells. The bacteria are also useful for producing molecule e.g.
 CC quantity or quality of protein or non-protein therapeutic molecule.
 CC Penicillin G, Erythromycin and Clavulanic acid. By biotransformation
 CC dominant negative alleles of the MMR gene are useful for producing higher
 CC quantities of recombinant polypeptides. The present sequence represents
 CC a yeast MHL1 protein.
 CC
 SQ Sequence 769 AA;
 QY Query: Match 29.7%; Score 1101; DB 22; Length 769;
 DB Best local similarity 32.8%; Pred. No. 6, 8e-82;
 DB Matches 265; Conservative 157; Mismatches 243; Indels 144; Gaps 19;
 QY 17 RIRLEESVNRIRIAGEVYORPSSAVKELIENSIGASSVSAVADGKLQVSDDH 76
 DB 4 RIKALDASVNVKIAAGELLISPVNALEKMMNSIDANATMIDLYVEGGIKVLOTDNCS 63
 QY 77 GIREEDLALICEBHRTSKLSAYEDLOTIKSGFGEALASMTYGVHYVYITTEGOLHEV 136
 DB 64 GINKADLPILCEHRTSKLSAYEDLOTIKSGFGEALASMTYGVHYVYITTEGOLHEV 123
 QY 137 RVSYRDGVMEVNEPPCAAVKGTQVWVNFYNNVARKKTLQNSNDYPRIVETISRF 196
 DB 124 RVSYAGKMLLESPKPVAGKDGTTILVEDLFNIRISRLRALRSHDEYSKILDVAGYALH 183
 QY 197 HINWTFESCRKIGANRADVHASTSSRLDAIRSYGASVVDLEIKYSEDADNDFE 255
 DB 184 SKDIGSCCKFGDSNLSVKPSYVQDRIRTFENKSVASMLITFHSKEDL--NLESV 241
 QY 256 DGYISNANYAKK-TIMLIFINDRLVDCALKRAIEFYVSAITLPOASKPEIYSIHL 314
 DB 242 DGKVCNLFISKSIISLIFINNRLVTCDLRRALNSYNTLPGRFPYILGIVDPA 301
 QY 315 HDVNIHPKKEVSLNOERIEETIRNAIEEKLMSNTTRIFQOALNSGIAQANP 374
 DB 302 AVDVNHPKREVRFLSODLIEKINOLHAEISALIDSRTFKASSISTNNPESILIPND 361
 QY 375 KV-----SEASMGSGTSQKIPVSAWVTRDP--RNPGRILTYWHGQSSNL 418
 DB 362 TIESDRNRKSLRQAVVENSYTTANSOLRRAKQENKVLVRI--SQAKITSPFLSSQOF 419
 QY 412 HGQSSNLEKKFDLVSVNVRSSRNQKADGLSSRHLLIVEIDSSFPGLDIYKNCYVGLADEAF 478
 DB 420 NFGSSTRKOLSEKRVINVSHSGAEKTLINESPQDANTINDNDLKDPKKKOLG 479

QY 443 -----LSSRHELVLEIDSFRHGLDIYKNCYV 471
 CC 480 KVSIADEKRNALPISMDGYIRVPEKERVANVLSIKKLRKVDSDIHRRELDIFANLVNY 539
 CC 472 GLADE--AFALIOHNRILYVNVVNSKELMYOALCRFENFNAIQSEPAQOELLVMA 529
 CC 540 GYDEERRRLAIIQHDKFLIDIGSVCELYEYIGLDFANFKINQSTNVSDDIVLYN 599
 CC 530 L--KDELMSEKDEKLELAENVTEILKENAMINEYFSIHIDOG-----KLTRL 579
 CC 600 LLSFDEL--NDASKERI--ISKI-----WDMSSMLNEYISIELVNDGLDNDLKSVLKSL 652
 CC 580 PVLVDOTYTPMDRLPEFVALAGNDVTWDEKECFRTVASAVGNFYALHPILPNPSONGI 639
 CC 653 PLLKGYISPLVLYKPEFYIRLGRKEDWEDQECIDGILREIALLY-----IP----- 699
 CC 640 HLYKKNDSMADEHAENDLISDENDVOELLAELAAMOREMTI-----OHVLEPSSRLF 695
 CC 700 -----DMVPKVDOTLDASLSEDEKAFINRKEHISLSLEHVLEPCKIKRR 742
 CC 696 LKPPKSMATDGTFOVASLEKLYIFERC 724
 CC 743 FLAPRHILKD--VVEIANTLPDLKVEFERC 769

RESULT 15

AAG63951
 ID AAG63951 standard; protein; 769 AA.

AC AAG63951;

DT 29-OCT-2001 (first entry)

DE Amino acid sequence of yeast mismatch repair protein MLH1.

XX MLH1, mismatch repair gene; MMR gene; hypermutable yeast.

OS Unidentified.

XX W0200162945-A1.

XX 30-ANG-2001.

XX 21-FEB-2001; 2001WO-US05447.

XX 23-FEB-2000; 2000US-0184336.

XX (UYJO) UNTY JOHNS HOPKINS.

XX (NICO/) NICOLAIDES N C.

XX (SASS/) SASS P M.

XX (GRAS/) GRASSO L.

XX (VOGE/) VOGELSTEIN B.

XX (KINZ/) KINZLER K W.

XX Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW;

XX WPI; 2001-522820/57.

XX N-PSDB; AAH75039.

XX Making hypermutable yeast that exhibit novel selected output traits for

XX commercial applications, comprises introducing polynucleotide

XX containing dominant negative allele of mismatch repair gene -

XX Disclosure; Page 36; 60pp; English.

CC yeast is useful for production, biocatalysis, bioremediation and drug
 CC discovery. It is also useful in genetic screens for the direct selection
 CC of variant subclones that exhibit new output traits. The hypermutable
 CC yeast is also useful in the manufacturing industry for the generation
 CC of new biochemicals, for detoxifying noxious chemicals from by-products
 CC of manufacturing processes or those used as catalysts, for remediation
 CC of toxins present in the environment including polychlorobenzenes, heavy
 CC metals and other environmental hazards for which there is a need to
 CC remove them from the environment. The yeast is also useful for producing
 CC increased quantity or quality of protein or non-protein therapeutic
 CC molecule e.g., Penicillin G, Erythromycin and Clavulanic acid, by
 CC biotransformation.

XX Sequence 769 AA;

Query Match 29.7%; Score 1101; DB 22; Length 769;

Best Local Similarity 32.8%; Pred. No. 6.8e-82;

Matches 265; Conservative 157; Mismatches 243; Indels 144; Gaps 19;

QY 17 RIRLEESVNRNRIAGEVIGTORPSSAVKELIENSIDAGASSVAVKDGKLIQVSDGH 76

DB 4 RIKALDSAVYVNTKIAAGEIILISPVNALEKEMENSIDANATMIDLYKGGIKVQITDNGS 63

QY 77 GIREDELALICERHTTSKLSAYEDLOTIKSMGEREALASMTYGVHTVTITTEGOLHGY 136

DB 64 GINKKADIPILCEHRTTSKLOKFEELSQIQTYGFEALASISHAVRYVTTKYEDRCAN 123

QY 137 RVASRGVMEVNEPRCAVAGCTOVMEVNLFFYNVARKKTIONSDOYKIVDFISRAVH 196

DB 124 RVSTAEKMLSEPPKPAVGKDGTTILVEDLFENFRLRALNSHDESKYLLDVGRAVH 183

QY 197 HINVFESCRKHGANRADVHASSTSRDLAIRSYGASVVDLEIKYS--YEDADSIFKM 255

DB 184 SKDIGFCKKFGDSNVSLSVAPSTYQDRIYFNKNSVANSNLTFHISKYEDL--NLESV 241

QY 256 DGYISNANYAKK--TIMILINDRLVDCIALKRAIEVYNTLPQASKPTIYSIHLPSE 314

DB 242 DGKVCNLFNFKSISLIFEFINRLVYCDLLRRLANSVSNYLPKGRPFYIYIVDPA 301

QY 315 HVDVNIHPTKEVSLNQEERIEFIRNAIEBKLNSTTRTFQFOQANLGGIAQANPKD 374

DB 302 AVDVNHPTRKREVFSLQDEILIEKIANQHAELSAIDTSRTFKASSISTPKESLIPND 361

QY 375 KV-----SEASMGSGTKSOKIPVSOVTRDPNPSGRLLTY-----W 411

DB 362 TIESDRNRKSLRQAVVENSVTYTNSQLRKAQENKLVIRDA--SQAKITPSFLSSQOF 419

QY 412 HGSSNLEKKFELVSVNVVRSRRNOK-----DA-----GD-- 442

DB 420 NFEGSSSTRQLSEPKVTVNSHSDAEKLTNSESOPDANTINDNDLKDPKKOKLGDY 479

QY 443 -----LSSRHELVLEIDSFRHGLDIYKNCYV 471

DB 480 KVSIADEKRNALPISMDGYIRVPEKERVANVLSIKKLRKVDSDIHRRELDIFANLVNY 539

QY 472 GLADE--AFALIOHNRILYVNVVNSKELMYOALCRFENFNAIQSEPAQOELLVMA 529

DB 540 GYDEERRRLAIIQHDKFLIDIGSVCELYEYIGLDFANFKINQSTNVSDDIVLYN 599

QY 530 L--KDELMSEKDEKLELAENVTEILKENAMINEYFSIHIDOG-----KLTRL 579

DB 600 LLSFDEL--NDASKERI--ISKI-----WDMSSMLNEYISIELVNDGLDNDLKSVLKSL 652

QY 580 PVLVDOTYTPMDRLPEFVALAGNDVTWDEKECFRTVASAVGNFYALHPILPNPSONGI 639

DB 653 PLLKGYISPLVLYKPEFYIRLGRKEDWEDQECIDGILREIALLY-----IP----- 699

QY 640 HLYKKNDSMADEHAENDLISDENDVOELLAELAAMOREMTI-----OHVLEPSSRLF 695

DB 700 -----DMVPKVDOTLDASLSEDEKAFINRKEHISLSLEHVLEPCKIKRR 742

QY 696 LKPPKSMATDGTFOVASLEKLYIFERC 724

Mon Apr 7 09:23:53 2003

DB 743 FIAPRIKD--VEIINLPDKVEEC 769

Search completed: March 27, 2003, 15:43:57
JOB time : 44 secs

us-09-954-950-2.1ag

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 15:45:12; Search time 36 Seconds

(without alignments)
1180,998 Million cell updates/sec

Title: US-09-954-950-2

Perfect score: 3709
Sequence: 1 MDEPSPRGCGACGEPPIR.....DGFYQVASELKLKIFERRC 724

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCF_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/CTOS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2305	67.5	737	10	US-09-749-601A-9
2	1453	39.2	756	9	US-10-079-429-2
3	1452	39.1	756	10	US-09-788-657-20
4	1101	29.7	769	10	US-09-788-657-15
5	599	10.1	669	10	US-09-815-242-5262
6	599	10.1	669	10	US-09-815-242-12334
7	565	13.2	649	10	US-09-815-242-13206
8	565	13.2	650	10	US-09-815-242-13667
9	547.5	14.8	629	10	US-09-815-242-10961
10	536	14.5	710	10	US-09-815-242-10895
11	471.5	12.7	633	10	US-09-815-242-12077
12	447.5	12.1	862	9	US-10-079-429-6
13	447.5	12.1	862	10	US-09-749-601A-11
14	403.5	10.9	859	10	US-09-788-657-16
15	402	10.8	779	9	US-09-749-601A-12
16	375	10.1	932	10	US-10-079-429-4
17	375	10.1	932	10	US-09-788-657-17
18	375	10.1	932	10	US-09-788-657-18
19	277.5	7.5	1151	10	US-09-749-601A-10

20	261	7.0	389	10	US-09-788-657-23	Sequence 23, Appl
21	249.5	6.7	133	10	US-09-749-601A-13	Sequence 13, Appl
22	246.5	6.6	147	10	US-09-749-601A-14	Sequence 14, Appl
23	232	6.3	133	10	US-09-788-657-21	Sequence 21, Appl
24	171	4.6	183	10	US-09-864-761-49029	Sequence 24, Appl
25	154.5	4.2	264	10	US-09-788-657-24	Sequence 24, Appl
26	154.5	4.2	264	10	US-09-788-657-25	Sequence 25, Appl
27	140.5	3.8	941	9	US-09-952-267-9	Sequence 9, Appl
28	137	3.7	1369	9	US-09-729-674-42	Sequence 42, Appl
29	128.5	3.5	831	9	US-09-952-267-1	Sequence 1, Appl
30	126.5	3.4	617	10	US-09-815-242-5762	Sequence 5762, Ap
31	126.5	3.4	2167	10	US-09-801-368-56	Sequence 56, Appl
32	124.5	3.4	77	10	US-09-925-300-1880	Sequence 1880, Ap
33	121.5	3.3	978	9	US-09-864-761-43207	Sequence 43207, A
34	121.5	3.3	3788	9	US-09-952-267-76	Sequence 76, Appl
35	120.5	3.2	457	10	US-09-815-242-12670	Sequence 12670, A
36	120.5	3.2	701	10	US-09-815-242-13002	Sequence 13002, A
37	120	3.2	2478	10	US-09-815-242-5816	Sequence 5816, Ap
38	120	3.2	2478	10	US-09-815-242-12967	Sequence 12967, A
39	118.5	3.2	1167	10	US-09-815-242-11522	Sequence 11522, A
40	116.5	3.1	3092	10	US-09-801-368-172	Sequence 172, App
41	116	3.1	873	9	US-09-952-267-13	Sequence 13, Appl
42	113.5	3.1	1786	9	US-09-742-096-3	Sequence 3, Appl
43	112	3.0	1332	10	US-09-982-091A-4	Sequence 4, Appl
44	112	3.0	2368	10	US-09-815-242-5635	Sequence 5635, Ap
45	112	3.0	2368	10	US-09-815-242-12389	Sequence 12389, A

ALIGNMENTS

```
RESULT 1
US-09-749-601A-9
Sequence 9, Application US/09749601A
Patent No. US20020128460A1
GENERAL INFORMATION:
APPLICANT: Nicolaides, Nicholas
APPLICANT: Grasso, Luigi
APPLICANT: Sass, Philip
APPLICANT: Kinzler, Kenneth
APPLICANT: Vogelstein, Bert
TITLE OF INVENTION: A method for generating hypermutable
FILE REFERENCE: 01107.00069
CURRENT APPLICATION NUMBER: US/09/749,601A
CURRENT FILING DATE: 2000-12-28
PRIOR APPLICATION NUMBER: 60/183,333
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 14
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 737
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-749-601A-9
Query Match 67.5%; Score 2305; DB 10; Length 737;
Best Local Similarity 66.4%; Pred. No. 7.6e-186;
Matches 482; Conservative 111; Mismatches 129; Indels 4; Gaps 3;
QY 2 DEPSPRGCGACGEPPIRLEESVNNRIAGEVYORPSSAVKELIENSIDAGASSVAV 61
DB 13 EEPSPATTIYPRBPPIKRIEESVNNRIAGEVYORPSSAVKELIENSIDAGASSVAV 72
QY 62 KDGGLKLIQVSDGHRFRFDALILCEHRTSTSLSAVEDIQTKSGFREALASMTYV 121
DB 73 KDGGLKLIQVSDGHRFRFDALILCEHRTSTSLSAVEDIQTKSGFREALASMTYV 132
QY 122 HAVVTITTEGOLHGYRYSYDGVNENPKPCAIVKGTQVWENLFYVMAKRTTIONSND 181
DB 133 HAVVTITTEGOLHGYRYSYDGVNENPKPCAIVKGTQVWENLFYVMAKRTTIONSND 192
QY 182 DYPRKIVDEISRFVAVHHINTVFSCRRKGANRADYHASSTSSRLDAIRSVYGVAVRDLIEI 241
```

Db 193 DYKTIYDLISMAIHHNVNSVSCRHGVAKVDVSVSPSLDSIRSVYGVAVKAKMLKV 252
OY 242 KVSIEDAADSIFFKADGYISANANYAKKTTMLFINDRLVDTALAKRAIEFYSATLPQAS 301
Db 253 EVSSDSSGCTFDEGEFISNSNYAKKTIILVFINDRIVECSAKRAIEIYATLPQAS 312
OY 302 KPEIYMSHLESEHVDVNIHPTKREVSLLNOERIIETIRNAIEEKIMSNTRIFOTQAL 361
Db 313 KPEIYMSHLESEHVDVNIHPTKREVSLLNOERIIETIRNAIEEKIMSNTRIFOTQAL 372
OY 362 NLSGIAQANPKDKVSEASMGSGTKSQKIPVSOVWRIDPBNPSCGRHTYWHGQSSNLEKK 421
Db 373 EYIQ-STLSQKSDSPVQSKPSGQKTCVYVKNKVRITSDSPAGRLHAFLOPKPQSLPDK 431
OY 422 FDLVS-VANVYASRRNOADADLSRHLVLEIDSSFRPLGDIYKNTCTYVGLADEARAL 480
Db 432 VSSLSVYRVSVRORNPKEIADLSSVOELLVGVSCCHPGLMLETYRNCTYVGMADVPAL 491
OY 481 IOHNTRLVNVVNIISKELTQOALCRFGNFNAIOLSEAPLOELLVAKDDEL--MSD 538
Db 492 VOYNHLLANVNIISKELTQOALCRFGNFNAIOLSEAPLOELLVAKDDEL--MSD 551
OY 539 EKDEKLEIAEVNTEILKENEMINEYFSIHIDOGKLTRLPVVLDQYTPMDRLPEVL 598
Db 552 TKDLEKLEIAEVNTEILKENEMINEYFSIHIDOGKLTRLPVVLDQYTPMDRLPEVL 611
OY 599 ALGNVYVWDEDECECRTRYASAVNGFYALHPLPBNPSCNGIHLKYKNDSDADEHAENDL 658
Db 612 CLGNVVEDEDECECRTRYASAVNGFYALHPLPBNPSCNGIHLKYKNDSDADEHAENDL 671
OY 659 ISDENVDQOELLAEAAAOREWITIOHVLFPMSRLFLKPKSMATDGTFOVASLEKLY 718
Db 672 VDMENNDQDILSDAENMAAOREWISIOHVLFPMSRLFLKPKSMATDGTFOVASLEKLY 731
OY 719 KIFERC 724
Db 732 KIFERC 737

RESULT 2
US-10-079-429-2
Sequence 2, Application US/10079429
Publication No. US20030027177A1
GENERAL INFORMATION:
APPLICANT: Haseltine et al.
TITLE OF INVENTION: Human DNA Mismatch Repair Proteins
FILE REFERENCE: PFI06P3D1
CURRENT FILING DATE: 2002-02-22
PRIOR APPLICATION NUMBER: PCT/US95/01035
PRIOR FILING DATE: 1995-01-25
PRIOR APPLICATION NUMBER: 08/466,024
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: 08/465,769
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: 08/294,312
PRIOR FILING DATE: 1994-08-23
PRIOR APPLICATION NUMBER: 08/210,143
PRIOR FILING DATE: 1994-03-16
PRIOR APPLICATION NUMBER: 08/187,757
PRIOR FILING DATE: 1994-01-27
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 756
TYPE: PRT
ORGANISM: homo sapiens
US-10-079-429-2

Query Match 39.2%, Score 1453, DB 9, Length 756;
Best Local Similarity 39.7%, Pred. No. 2, 7e-104;
Matches 312; Conservative 144; Mismatches 214; Indels 116; Gaps 11;

OY 18 IRRLEESVNRNIAAGEYIOPRSSAVKELIENSIDAGASSVAVAKDGLKLIQVSDGHC 77
Db 8 IRRLEESVNRNIAAGEYIOPRSSAVKELIENSIDAGASSVAVAKDGLKLIQVSDGHC 67
OY 78 IRFEDLAILCERHTTSKLSAVEDLOTKSMCFGEALASMTYGVHTVTTTTSQOLHGR 137
Db 68 IRKEDLILVCERHTTSKLSAVEDLOTKSMCFGEALASMTYGVHTVTTTTSQOLHGR 127
OY 138 VSYRSDGVNENPKPCAVKGTQVAVENLFFNMVARKKTLIONSDYKIDYFISPAVH 197
Db 128 ASYSDGLKAPKPCAGNQTQITVEDLFTYIATRRALKNPSEYKILIEVGRYVHN 187
OY 198 INVFSCRKHSANRADVHSASTSRDLAIRSVYASVARDLIEIKVSEDAADSIFFKMG 257
Db 188 AGISEFVAKGGEIYADVADRTLPNASTVDNINRSEFNANVRELIEICGDKTLA--FMMG 244
OY 258 YISNANYAKKTIILFINDRLVDTALAKRAIEFYSATLPQASKPTIYMSHLPSEHVD 317
Db 245 YISNANYAKKTIILFINDRLVDTALAKRAIEFYSATLPQASKPTIYMSHLPSEHVD 304
OY 318 VNIHPTKREVSLLNOERIIETIRNAIEEKIMSNTRIFOTQALNSGIAQANPQ--KDK 375
Db 305 VNIHPTKREVSLLNOERIIETIRNAIEEKIMSNTRIFOTQALNSGIAQANPQ--KDK 363
OY 376 VSEASMGSGTKSQKIPVSOVWRIDPBNPSCGRHTYWHGQSSNLEKK 426
Db 364 TSLSSSTSSGSDVYVAKQVWRTDSRQ--KLDAFIQPLSKPLSSQOAVITEKTDISS 421
OY 427 VRNVYR-----SRNOKDAGDS----- 444
Db 422 GRARODEMELDAPAEVAAKNQSLBDDTTKTSSEMSKRGPTSSNPKRRHREDSDVEM 481
OY 445 -----SRHELIV-----EIDSSFRHGLDIYKNCYVGLADEAPALI 481
Db 482 VEDSRKMTACTPARRKIINLTSVLSQEBINQGHVLAEMLNHSFVCCVAPQNALA 541
OY 542 OHOFKLYLLNTKLSSELFYQIILYDFANFGVLRSEAPFLDAMLALDPSGAWTEED 601
Db 542 DEKLEIAEVNTEILKENEMINEYFSIHIDOGKLTRLPVVLDQYTPMDRLPEVLALG 601
OY 602 GKPEGLAIYIEFLKKNKEMADYFSEIDEGNIGLPLIDNVPVLEGLPFTILLA 661
Db 602 NDVFWDDKECEFRVYASAVNGFYALHPLPBNPSCNGIHLKYKNDSDADEHAENDLSD 661
OY 662 TEVWDEDECEFRVYASAVNGFYALHPLPBNPSCNGIHLKYKNDSDADEHAENDLSD 693
Db 662 ENDV---DOELLAEAAAOREWITIOHVLFPMSRLFLKPKSMATDGTFOVASLEKLY 718
OY 694 ESTLSGQSEVPGSIPNSM---KMTVEIYVYKALRSHILPKHFTEDGNITLNLNPLLY 750
Db 719 KIFERC 724
Db 751 KIFERC 756

RESULT 3
US-08-788-657-20
Sequence 20, Application US/09788657
Patent No. US20020123149A1
GENERAL INFORMATION:
APPLICANT: Sasse, Philip
APPLICANT: Nicolaides, Nicholas
APPLICANT: Kinzler, Kenneth
APPLICANT: Grasso, Luigi
APPLICANT: Vogelstein, Bert
TITLE OF INVENTION: Methods for generating hypermutable
FILE REFERENCE: 01107.00097
CURRENT APPLICATION NUMBER: US/09/788,657
CURRENT FILING DATE: 2001-02-21

PRIOR APPLICATION NUMBER: 60/184,336
PRIOR FILING DATE: 2000-02-25
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 20
LENGTH: 756
TYPE: prt
ORGANISM: Homo sapiens
US-09-788-657-20

Query Match 39.1%; Score 1452; DB 10; Length 756;
Best Local Similarity 39.6%; Pred. No. 3, 2e-104;
Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;

18 IRRLEESVNRITAGEVIRPSSAVKELIENSIDAGASSVAVKGGKLIQVSDGHC 77
18 IRRLEESVNRITAGEVIRPSSAVKELIENSIDAGASSVAVKGGKLIQVSDGHC 77
8 IRRLEESVNRITAGEVIRPSSAVKELIENSIDAGASSVAVKGGKLIQVSDGHC 67
78 IRRLEESVNRITAGEVIRPSSAVKELIENSIDAGASSVAVKGGKLIQVSDGHC 137
68 IRRLEESVNRITAGEVIRPSSAVKELIENSIDAGASSVAVKGGKLIQVSDGHC 127
138 VSYRQGVMEKNEPRCAAVAGTQVMEVNFYNNVARKKTLONSDDPKIYDFISRAVHH 197
128 ASYSDCKLAPKPCAGNOGTQITVEDLYNTATRRKALKNSPEEYKILIEVGRYSVHN 187
198 INVETSCRRHGANRADVHASTSSRLDAIRSVYASVVRDLIEIKVSEDAADSIKFMKG 257
188 AGISFVKKOGETVADVRLPNASTVDNIRSGINAVSRLIEIGEDKTLA---FKMG 244
258 YISNANVAKKTIIMLIFINDRLVDCATKRAIEFYSAIIPQASKPFIYSHLPSEHVD 317
245 YISNANVAKKTIIMLIFINDRLVDCATKRAIEFYSAIIPQASKPFIYSHLPSEHVD 304
318 VNIHPKKEVSLNOERIIETIRNAIEKILMSNTTRIFOTQALNLSGIAOAMPD--KDK 375
305 VANHPKHEVHLFHEESIEERVOQHIESKILGSSNMRTYOTL-LPLGLGPGSEKWKST 363
376 VSEASGSGTSSOKIPVSGMVRTDPRNPSGRLLTYHNGOSSNL-----EKKEDLVS 426
364 TSITSSTSGSSDKVYAHQMVRTDSREQ--KIDAFIQLPSKPLSQPQAVTEPDKTDISS 421
427 VRRVVR-----SRNQKADGLS-----444
422 GRARODEEMLELPAPAEVAAKNQSLGDTTKGTSEMSEKRGPTSSNPKRRHREDSVEM 481
445 -----SRNELLY-----EIDSPHGLDIYKNCYVGLADEAPALI 481
482 VEDDSKKEMTACTPRRIINLTSTVLSLOEINQGHVEVLRKMLHNSFVGCYNPQWALA 541
482 OHNTRLYLVNANNISKELMYOALCFRGNFNAIOLSEPAILOELVMAKDELMSDEKD 541
542 OHNTRLYLVNANNISKELMYOALCFRGNFNAIOLSEPAILOELVMAKDELMSDEKD 541
542 OHNTRLYLVNANNISKELMYOALCFRGNFNAIOLSEPAILOELVMAKDELMSDEKD 541
542 DELELAENVTELKEMAKINEYESIHDQCKLRLPYVLDQYTPDMDLPEFVALG 601
602 GPEGGAETIYELVLRKKAEMADYFLEIDEBGNLGLPDLIDNYVPLEGIFILIRLA 661
602 NDVTWDEKCEFTVASAVGNFYALHPILPNPNSGNGIHLKKNRSMADENHENDLISD 661
662 TEYNMDEKECFESLKECAMFYST-----RKQYISE 663
662 ENDV---DDELLAEBAANAQREMTIQHVLFPMSRLFLRPKSMADGTFVQVASELEKLY 718
694 ESTLSQSGSEVPISPSNSW---KMTVEHLYKALRSHILPDKHFTDGNILQALNPDLV 750
719 KIFERC 724
751 KIFERC 756

RESULT 4
US-09-788-657-15

Sequence 15, Application US/09788657
Patent No. US20020123149A1
GENERAL INFORMATION:
APPLICANT: Nicolaides, Nicholas
APPLICANT: Sassi, Philip
APPLICANT: Kinzler, Kenneth
APPLICANT: Grasso, Luigi
APPLICANT: Vogelstein, Bert
TITLE OF INVENTION: Methods for generating hypermutable
FILE REFERENCE: 01107.00097
CURRENT APPLICATION NUMBER: US/09/788,657
PRIOR APPLICATION NUMBER: 60/184,336
PRIOR FILING DATE: 2000-02-23
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 15
LENGTH: 769
TYPE: prt
ORGANISM: Saccharomyces cerevisiae
US-09-788-657-15

Query Match 29.7%; Score 1101; DB 10; Length 769;
Best Local Similarity 32.8%; Pred. No. 5, 2e-77;
Matches 265; Conservative 157; Mismatches 243; Indels 144; Gaps 19;

17 IRRLEESVNRITAGEVIRPSSAVKELIENSIDAGASSVAVKGGKLIQVSDGHC 76
4 RIKALDSVYNNKTAAGEIISPVNALEKEMENSIDANATMIDILVEGGIKVQIDTNGS 63
77 GIREFALICERHTTSKLSAYEDLOTIKSMGREGALASMTYVGHVTTTTEGOLHGY 136
64 GINKADPLICERHTTSKLOKPFEDLSQIOTYGERGELASISHVARTVTTTKEDCAM 123
137 RVSYRQGVMEKNEPRCAAVAGTQVMEVNFYNNVARKKTLONSDDPKIYDFISRAVHH 196
124 RVSYRQGVMEKNEPRCAAVAGTQVMEVNFYNNVARKKTLONSDDPKIYDFISRAVHH 183
197 HINVTSCRRHGANRADVHASTSSRLDAIRSVYASVVRDLIEIKVSEDAADSIKFMKG 255
184 SKDIGFCKKFGGNSYSLVSKPSTVODRIRTVFNKSVASNLTFHISKVEDL--NLESV 241
256 DGYISNANVAKK--ITMILFINDRLVDCATKRAIEFYSAIIPQASKPFIYSHLPSE 314
242 DGRVONLNFISKSSISLIFINNRLVTCDLRRALNSVYNYLPKGRPFITIGIYDPA 301
315 HDVNTHPTKKEVSLNOERIIETIRNAIEKILMSNTTRIFOTQALNLSGIAOAMPD 374
302 AVDVNHPKHEVHLFHEESIEERVOQHIESKILGSSNMRTYOTL-LPLGLGPGSEKWKST 363
375 KV-----SEASGSGTSSOKIPVSGMVRTDPRNPSGRLLTY-----W 411
362 TIESDRNRKSLROAQVENSITTANSOLRAKROENKLVIDA--SOAKITPSSSSQOF 419
412 HGSSNLEKFFDVSVANNVRSRRNK-----DA-----GD--442
420 NFEGSSSTRQLSPKVTYNVSHQEAELTLNESBPQADANTINDNDLKQPKKKQKGDY 479
443 -----LSRHELVLEIDSSFRPGLDIYKNCYV 471
480 KVPISADDEKNALPISKGYIRVPERVNNVNLISIKKLRKVDYSIRELTDIFANLNVY 539
472 GLADE--AFALIQHNPRLYLVNANNISKELMYOALCFRGNFNAIOLSEPAILOELVMA 529
540 GYVDEERRRLAIGHDLKFLIDYGSVCYELFTQIGLDFRNFKINLOSTNSDDIYLVN 599
530 L--KDELMSEKDEKLELAENVTELKEMAKINEYESIHDQCKLRLPYVLDQYTPDMDLPEFVALG 579
600 LSEFDEL--NDASKERI--ISKI-----WDMSSMLNLEYSTIELVNOGLDNDLKSVKLSL 652
580 PVVLDQYTPDMDLPEFVALGNDVYTWDEKCEFTVASAVGNFYALHPILPNPNSGNGI 639

Db 653 PULLKTIPLVLPFFITRLGKEVMEDEQELDGLIREALLY-----IP----- 699
QY 640 HLYKKNRDSMADEHAEENDLISDENVDQELIAEAAQOREWTI-----OHVLEPQMRL 695
Db 700 -----DWPVKYDTLDASLSEDEKQFIRKHHISLLEHVLPFCIKR 742
QY 696 LRPKSMADGTFVQVASELKYIFERC 724
Db 743 FLAPRHILD--VEELANPLDYKVEFC 769

RESULT 5
US-09-815-242-5262
Sequence 5262, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815.242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5262
LENGTH: 669
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5262

Query Match 16.1%; Score 599; DB 10; Length 669;
Best Local Similarity 27.2%; Pred. No. 3.4e-38;
Matches 179; Conservative 115; Mismatches 203; Indels 160; Gaps 19;

QY 17 RIRLEESVNRVRIAGEVYIORPSSAVKELIENSIDAGASSSVANKOGKILIQVSDG 76
Db 3 KIKELQTLAKIAGEVERPSSVKKLEENALDAGATEISIEVEGSGVIRVWNGS 62
QY 77 GIREEDLALICERTTSLASVEDLOTIKSGFGEALASTYVGHVVTITTEGOLHG 136
Db 63 GIEADGLVHRHATSLDDEDLPHRTIGFGEALASISSVAKVTLKCTD--NANGN 121
QY 137 RVRTRDGVMEPPKCAVKTQVWENTFTNMVARKKTLONSDDYPKIVDFISRPVH 196
Db 122 EIVYENGELIHH--KPAKAKCTDILVESLFYNTPARLKIKSLVTELGKLTIDIVNRAMS 180
QY 197 HINTFOSRKGANRADVHASTSSRLDAISVYGASVNRRLIETIKSYEAAASIRKMD 256
Db 181 HPDIRIALISGKMTSLNDSGRIN--EVAETIGMVARDLHIS--GDTSD--YHIE 233
QY 257 GYI-----SNANYVAKKITMILFINDRLVDCALKRAIFVYSATLPQASKRPFIMSI 309

Db 234 GFVAKPEHSRKNKHTIS-----IFINGRIKNEKMLAKILEGYPHLITIGRPICINYI 287
QY 310 HLPSEVVDVNIHTKVEGLVNOERITETIRNAIEKILMSNTTRIFQOAL-----NLSG 365
Db 288 EMDPILVADVNVHPTKLEVALSKEEQIYOLIVSKIOE-----AFKDRILIPKNNIDY 338
QY 366 IAOANPOKDVSEASWGSGTSOKIPVSQWRTDNRNPSGRILTYWHQSSNLEKFPOLV 425
Db 339 V-----PKKKVLS-----FEOKIEFEORONTE-----NNOETFSSE 373
QY 426 SVRNVVRSRNOKD-----AGDLS-----RHEL 449
Db 374 ESNKRPFAENONDEIVIKEDSYNPFVTKTSLLADDESSYNNTRKDEDFPKKOEI 433
QY 450 LVEIDSSPHGLDLYKN-----CTYGLADEAPAL 480
Db 434 LDEMOTFDSNDTSVQNEKASDDYDVNDIKGTSKDPKRRIPYMEIYQVHGTYII 493
QY 481 IOHNTRLVNVVNTSKELMYOALCRFNFENAIQLEBPAPLOELVMAL-----KDEELM 536
Db 494 AQNEFGMTMIDQHAQERIKYTFPDKIGEYI-----NEYODLLPLTFHFSKDEQLV 546
QY 537 SPEKDEKLEIA-----EVTETILKENAEMINEYFSIHI 570
Db 547 IDQYKNEIQVGMLEHFGHDYIVSYVPWPKDEVE-EIKDMIELILEKKVDI 602

RESULT 6
US-09-815-242-12334
Sequence 12334, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815.242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12334
LENGTH: 669
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12334

Query Match 16.1%; Score 599; DB 10; Length 669;
Best Local Similarity 27.2%; Pred. No. 3.4e-38;
Matches 179; Conservative 115; Mismatches 203; Indels 160; Gaps 19;

QY 17 RIRLEESVNRVRIAGEVYIORPSSAVKELIENSIDAGASSSVANKOGKILIQVSDG 76

Db 3 KKEIOTSLANKIAGEVVERSSVYKELLENAIDAGATEISTEVEBSGOSIRVNDGS 62
Qy 77 GIREFEDIALICERHTTSKLSAYEDLOTIKSMGFRGALASMTYVGHVTVTTITTEGOLHGY 136
Db 63 GTEAEDLGIVFPHHATSKIDODEDLPHIRITLGRGALASISSAVATLKTICD-NANGN 121
Qy 137 RVSYRDGVNENEPKCAAVKGTQVAVENLFYNNVARKKTLONSNDYPRKIDFISPAVH 196
Db 122 EIVYENGELINH-KPAKAKKGTDILVESLFYNTPARLKIKSLYTELGKTTIDVNNMAMS 180
Qy 197 HINVFSCRHGANRADVSASTSSRLDAIRSYGASVVRDLIEIKVSEDADAIFKMD 256
Db 181 HEDIRIALISDQKTMJSTINGSGRTN-EVMAETYGKAVADLVHIS-GDTS-DYHIE 233
Qy 257 GYI-----SNANYAKKITMILFINDRVDTALKRAIEFYSAATLPQASKPTIYMSI 309
Db 234 GYAPKPEHRSRKHYS-----IFINGRYIKNFMLNKAILBEGYHLLTIGRPICITINI 287
Qy 310 HLPSEHVDVNIHPTKKEVSLNDRLEITIRNAIEEKLANSNTTRIFOTQAL---NLG 365
Db 288 EMDPILYDVNVHPTKLEVLRSKEEQLYOLIVSKIOE-----AFKDRILIPKNNLDY 338
Qy 366 IQANPOKXKVSASGSGTKSOKIPVQWVRDPNPSGRRLTYHMGSSNLEKKFDV 425
Db 339 V-----PKKKVLSH-----PEOKIEFEOQNT-----NNOEKTFSS 373
Qy 426 SVRNVRSHRNOKD-----AGDLS-----RHEL 449
Db 374 ESNKRFEMENONDELIVIKEDSYNPFVTKTSESLIADDESSGNNRREKDEDFKQOEL 433
Qy 450 LVEIDSSHPGLDLYKN-----CTYGLADBAFAL 480
Db 434 LQEMQPTFSDNDTSYQVYENKASDDYVDNDIKTKSKDPKRRIPIYMEIVGQVHGTYII 493
Qy 481 IOHNTFLYLVNVNISKELMYOALCRFGFNALIOLEAPLOELVMAL---KODELM 536
Db 494 AQEREGMIMIDQAAOERIKETFEKDIKIGVT-----NEVODILLPLTFHESKQEOV 546
Qy 537 SDEKDEKELEIA-----EVNTEILKENAMINEYSIH 570
Db 547 IQYKNEIQOVGIMLEHFGHDIYSSYPWPFKDEVE-EIKMDIELILEKKVDI 602

RESULT 7
US-09-815-242-13206
Sequence 13206, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308.
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO.13206
LENGTH: 649
TYPE: PR
ORGANISM: Streptococcus pneumoniae
US-09-815-242-13206

Query Match 15.2%; Score 565; DB 10; Length 649;
Best Local Similarity 27.7%; Pred. No.1,4e-35;
Matches 162; Conservative 115; Mismatches 219; Indels 88; Gaps 16;

Qy 18 IRLRESVVRIRIAGEVTEIOPSSAVKELINSIDAGSSVSVAVKDGLKLIQVSDGCH 77
Db 4 IIEPELANQIAGEVTEIOPSSAVKELVBNADAGSSQIIEIEGKLVKVOITTDGCH 63
Qy 78 IREFEDIALICERHTTSKLSAYEDLOTIKSMGFRGALASMTYVGHVTVTTITTEGOLHGY 137
Db 64 IADDEVIALRRHATSKIRKQADLFIRITLGRGALPSIASVVLTLAIVDAGASHGK 123
Qy 138 VSTRDGVNENEPKCAAVKGTQVAVENLFYNNVARKKTLONSNDYPRKIDFISPAVH 197
Db 124 LVARGEVE-EVLPATSPVGTKYCEDLFENTPARLKYSQAELSHIIDYVRLGLAH 182
Qy 198 INVTFSCRHGANRADVSASTSSRLDAIRSYGASVVRDLIEIKVSEDADAIFKMD 257
Db 183 PEISFSLISG--KEMTRTGOLKQALIGTGLVSAKMIELNS-----DIDFISG 235
Qy 258 YIS-----NANYAKKITMILFINDRVDTALKRAIEFYSAATLPQASKPTIYMSI 310
Db 236 FVSLPELTRANRNYIS-----LFINGRYIKNFLLNALIDGOSKLMVGRPLAVIHH 289
Qy 311 LPSEHVDVNIHPTKKEVSLNDRLEITIRNAIEE-----KLMNSVTR 354
Db 290 IDPYLADVNVHPTKOEIRIKELMTLVSEALANSIKETOLLPDALENLAKSTVRRBK 349
Qy 355 IFOT-----QALNLSQIAQANPOKXKVSASGSGTKSOKIPVQWVRDPNPSGRRL 408
Db 350 VEOTIILPKENTLYETETSPRSQTEVADYQVELDDEGDDLPAKETIDRLTKPAKLH 409
Qy 409 TYWHGSSNLEK---KFDLVSVRNVRSHRNOKDAGDLSRHELLVEIDSSHPGLDI 464
Db 410 -FAERKPRANYDQDHPRELDAI-----DKADKLEREASSPHEL--- 449
Qy 465 VKKCTYVGLADEAPALIOHNTFLYLVNVNISKELMYOALCRFGFNALIOLEAPLOE 524
Db 450 ---EFFGOMHGTYLFAQGRDGLYIIDQAAOERIKETFEKDIKIGVT-----QQ 498
Qy 525 LVMAL---KODELMSDEK---DDEKLEIAFY--NTEILKEN 558
Db 499 LVPYIPEFPADALRLKEKMPLEEVGLAEYGENOFILREH 542

RESULT 8
US-09-815-242-13667
Sequence 13667, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 13667
 LENGTH: 650
 TYPE: PRT
 ORGANISM: Streptococcus pneumoniae
 US-09-815-242-13667

Query Match 15.2%; Score 565; DB 10; Length 650;
 Best Local Similarity 27.7%; Pred. No. 1.4e-35;

Matches 162; Conservative 115; Mismatches 219; Indels 88; Gaps 16;

18 IRRLEESVYVNRVIAAGEVIGORPSSAVKELIENSLDAGASSVAVAKDGLIKLOVSDGCHG 77
 4 IIELEPMLANOIAAGEVERPASYVKELVENAIDAGSSGIIIEIEAGKKVQIINDNGHG 63
 78 IREFDLALCEHRTSKLSAYEDLOTISKMGFRGALASMTYVGVYVTTTTEGOLHGR 137
 64 IAHDEVELLRHAAVSKIRKNOADLEIRTLGFRGALPSIASVSVALTLTAADGASHGK 123
 138 VSYRDGVNENPKCAAVKQGVWVENLFYNNVARKKTLONSNDYPRIVDFISPAVHH 197
 124 LVARGGEVE-ETIPATSVGTVKCYEDLEFNTPARLTKMSQOALSHIIDVNLGLAH 182
 198 INWTECRKHGANRADVHASTSHLDAIRSVYASVVRDLIEIKVSEDAADSIFFKMDG 257
 183 PEISFSLSDG--KEMPTAGTGOLRAIAGIYGLVSAKKMEIRNS-----DLDFEISG 235
 258 YIS-----NANYAKKITMILFINDLVCTALAKRAIEVYSAITLQASKPFYMSIH 310
 236 FVSLPELTKRANKNIS-----LFTNGRYINFLNRAILDGFSGKMWGRPLAVIH 289
 311 LPSEHVDVNIHPTKKEVSLNQRIL--ETIRNAIEE-----KIMSNTR 354
 290 IDPYLADVNVHPTKQEVNISKEKELMTLVSEALNSLKEQTLIPDALNLAKSTYRNRK 349
 355 IFOR-----QALNLSGIAQANPOKDKYSEASMGSGTKSQKIPVSQWVRTDPRNPSGLH 408
 350 VEQTLPLKENTLTYEKTSPSPQTEVADYQVELTDEGQDLTLFAKTELLRLTKPAKHL 409
 409 TYWGGQSSNLEK---KFDLVSVRNVVSRNRKQADGLSRHSHLVEIDSSFFHGGLDI 464
 410 -FAEKRPANVQDLHPDLAST-----DAYDKLEEESSSPFEL-- 449
 465 VKNCTVGLADEAPALIQHNTRLVYNNVNSIKELMTQOALCRGFENAIOLSEAPLOE 524
 450 -----EFGGOMGTLYAAGRGDLYIIDQHAQEKVKEEYHESJGNVQDS-----QQ 498
 525 LLYVAL---KDDLEMSRK-----DDEKLETAEV--NTEIILKEN 558
 499 LTVPIEFPPADALRLKERMPILLEEVGVFLAEGENQFILREH 542

RESULT 9

US-09-815-242-10961

; Sequence 10961, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 10961
 LENGTH: 629
 TYPE: PRT
 ORGANISM: Haemophilus influenzae
 US-09-815-242-10961

Query Match 14.8%; Score 547.5; DB 10; Length 629;
 Best Local Similarity 27.3%; Pred. No. 3e-34;
 Matches 174; Conservative 119; Mismatches 251; Indels 93; Gaps 19;

18 IRRLEESVYVNRVIAAGEVIGORPSSAVKELIENSLDAGASSVAVAKDGLIKLOVSDGCHG 77
 3 IKILSPOLANOIAAGEVERPASYVKELVENAIDAGSSGIIIEIEAGKKVQIINDNGHG 62
 78 IREFDLALCEHRTSKLSAYEDLOTISKMGFRGALASMTYVGVYVTTTTEGOLHGR 137
 63 IPKEBLSLALRHATSKLADDDLEALISLGFREGALASISVGRITLSTFEQTEKMQ 122
 138 VSYRDGVNENPKCAAVKQGVWVENLFYNNVARKKTLONSNDYPRIVDFISPAVHH 197
 123 VYAGRDMEETIKPASHVGVGTVVEANLFNTPARRKLRTDKTEFSHIDEVIRIALTK 182
 198 INWTECRKHGANRADVHSA--STSSRLDAIRSVYASVVRDLIEIKVSEDAADSIFFKMD 256
 183 FNTAFETLHNCKIIRYRPAEAINOOLKRVANICGDDPVKALAIENKHD-----LHLS 237
 257 GYISNANYVAKKITM--ILFINDLVCTALAKRAIEVYSAITLQASKPFYMSIHLPSEH 315
 238 GVAATPMPNSRTQNDLSYCYINGRWVRKVISHAIRQAAQYLPDAYPAVLFIDLPMD 297
 316 VDVNIHPTKKEVSLNQRIL--ETIRNAIEEKLMSNTRTRIFQOALNLSGIAQANPOKDK 375
 298 VDVNVHPTKKEVSRHQOQLHDFIYEGISBALNNOEQIINMHTDS-----AVENHEBT 351
 376 VSE-----ASMGSGT-----SKQIPVSQWVRTDPRNPSGLRHTYWHGQSS 416
 352 VREQPMVNSIRPNRATAGQNSFAPOYHEKPOQNPSPHSNTPMPFNHNSGTGRDYRSAPPS 411
 417 NLEKKFDLVSVRNVSRNRKQADGLSSRH-----ELLVEIDSSFFHGGLDIYK 466
 412 KTEORLYAELELRTLPRA--QKDISDPAQONISDTAKISTEI---IECSHIALSLIE 466
 467 NCTVYGLADEAPALIQHNTRLVYNNVNSIKELMTQOALCRGFENAIOLSEAPLOEL 526
 467 N-----RALLLQONODEFLISTLEKLR-IOMQAL-----KOIQEQALLLP-I 509

QY 527 VALKDELMSDEK--DDEKLEIAEYNTLKKNAEMINEFSIHIDQDKLT--RLPVV 582
DB 510 VERLTESQOAMQOYSDDER---KIGFEFIENQAOI-----RLTLNVPSPA 552
QY 583 LDQYTPMDRLPEFVALANDVTWMDDEKQCFRIVASA 619
DB 553 L-----RTONLQKCVAM---LIRDENSSSFLTALCA 581

RESULT 10
US-09-815-242-10895
Sequence 10895, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlisen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT FILING DATE: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 10895
LENGTH: 710
TYPE: PRF
ORGANISM: Enterococcus faecalis
US-09-815-242-10895

Query Match 14.5%; Score 536; DB 10; Length 710;
Best Local Similarity 26.1%; Pred. No. 2,8e-33;
Matches 179; Conservative 113; Mismatches 243; Indels 152; Gaps 21;

QY 17 RIRLEESVNRRIAGEVIQRPSSAVKEIENSIDAGASSVAVKDGKILQVSDGH 76
DB 3 KIOELSEOLANOIAAGEVEVPASVYKELVENALDAGSQIDIFIEAGKTIQIDNGE 62
QY 77 GIREFDLALCERHTSKISAYEDLOTISKMGREGALASMTYGVHTYTTTIEGOLHG 136
DB 63 GAKKEVLAIAFRKHAITSKITHRDLEIRSLGREGALPSIASVSMTIETATAEEEGS 122
QY 137 RVSRYDGVANENEPKCAAVKGVOMENLFYNNVARKKTIONSNDYPRKVDISFPVAVH 196
DB 123 YILKKGAYE-ENRPAALAKKGTMYSNLFYNPALAKTYKTIOTELANGDIYNKLALS 181
QY 197 HINWTFSCRKHGANRADVHSA---STSSRLD---AIRSVYASVVRDLIEIKVSYEDA 249
DB 182 HPEVAFRL-----VHDGHRMSTGTNGDKOTIAGIYGISTAKMKIKI-----EGK 227
QY 250 DSIFKKNDGYIS-----NANTYAKKITMILFINDRLVDTCTAKRAIEFYYSITLQASK 302
DB 228 DUDFTLITGVISLPEVTIRASRNYS-----TIIINGRYIKNFALNKAIIVAGYGSKLWGRP 281

QY 303 PEIYMSIHDPSEHVDVNIHPFKREVSLNNOERILLETIRNAIEEKIMNSNTTRJFOQALN 362
DB 282 PIAVLEIEMOPLLYVDVNVHPTKQEVRLSKERKIDMALIEQAIOEVL-----SQEOLIPNADN 338
QY 363 LSGIAQANPKOKVASEMSKSGTSKOKTIPVSQWKRDPNP-SGRLLTYHAGGSSNLEKK 421
DB 339 LR-FKKRLPEPKAEQOEIPLST--EQPAVKORGSISYNPACGRFA----- 383
QY 422 FDLVSVRNVVRNRQKADGLSSRHLLVEIDSFPGLDIYKNTGYGLADEAFA-- 479
DB 384 -----EHANDERFPASSAPAVNOTERAVP-----HDDPPAPP 416
QY 480 -----LIQHTRLYLNVVNIISKELMYQALCRGFENAIQLSPAPLQELLYWALKDDE 534
DB 417 VPESTVAEETSAYETSAAVDSSENHETA-----SNEQOVIAEETKITE 461
QY 535 LMSDEKDEKLEIAEYNTLKKNAEMINEFSIHIDQDKLTRLPVLDQYTPMDRLP 594
DB 462 VPDS--PSLRSEADEVFOEEMKLPPEPDANSATSOBELQALNKLEEREP-TERFP 517
QY 595 EFVIALGNDVTWMDKEKCFRTVASAVGNFYALHPILPNPSNGIHLXKKNRDSM--ADE 652
DB 518 E-----LEYFCOMH-----GTYLPAQSKDGLFTIDQ 543
QY 653 HAENDLISDE-----NDVQDELL 670
DB 544 HAAQERIKKEYFREKIGEVSDDLQDELL 570

RESULT 11
US-09-815-242-12077
Sequence 12077, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlisen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 12077
LENGTH: 633
TYPE: PRF
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12077

Query Match 12.7%; Score 471.5; DB 10; Length 633;
Best Local Similarity 33.1%; Pred. No. 2,4e-28;
Matches 112; Conservative 66; Mismatches 149; Indels 11; Gaps 4;

OY 14 EPPRIREEVYNNRIAGEVIOBPSSAAVELIENSLDAGASSVAVKDGKILQVSD 73
 DB 3 EAPRIOLSPRIANOJAGEVEVERPASVAKELLENSLDAGSRIDIVEBOGGIKILRVSD 62
 OY 74 DGHGIRFEDLALICERHTSKLSAYEDLOTISMGPRGALASMTYGVHTTTEGOL 133
 DB 63 DGRGIPADDLALANHAHSKILREJEDLERVMSLGRGALASSISVAALTMISRIADAG 122
 OY 134 HGYRVSYRDNVEMNEPKPCAAYKGTQVWENLFYNNVARKTIONSNDYPRIVDEISRE 193
 DB 123 EAMQVETESGRDQMPVQPAHVGTSYVERDLPFNTPARAKRLRAKTEFDLQEVYIKL 182
 OY 194 AVHINVTFSCKHGANRADVHASTS-SRLDAIRSVYGASVVDLIEIKVSYEDAASI 252
 DB 183 ALARPDVAFHLHNGKTIFFALHEARDELARARVAGVCGOAFLEQALPFEVERNG----- 237
 OY 253 FEMDGYISNANYVAKKIM-ILFINDRLVDCALAKRAIEFYVSATLPQASKPIYMSIHL 311
 DB 238 LHLMGWGLPFTFSRQOPDLQYFYVNGRWVRDKLVAAVAAQARVDVLYNGRHPTFVLFEEV 297
 OY 312 PSEHVDVNIHPTKEV-----SLNQRRIETIRNAIEE 345
 DB 298 DPVAVDVVNHPTKHEVRERDRSRMVDLXGTLHRAIGE 335

RESULT 12

US-10-079-429-6
 : Sequence 6, Application US/10079429
 : Publication No. US20030027177A1
 : GENERAL INFORMATION:
 : APPLICANT: Haseltine et al.
 : TITLE OF INVENTION: Human DNA Mismatch Repair Proteins
 : FILE REFERENCE: PF106P3D1
 : CURRENT APPLICATION NUMBER: US/10/079, 429
 : PRIOR FILING DATE: 2002-02-22
 : PRIOR APPLICATION NUMBER: PCT/US95/01035
 : PRIOR FILING DATE: 1995-01-25
 : PRIOR APPLICATION NUMBER: 08/468, 024
 : PRIOR FILING DATE: 1995-06-06
 : PRIOR APPLICATION NUMBER: 08/465, 769
 : PRIOR FILING DATE: 1995-06-06
 : PRIOR APPLICATION NUMBER: 08/294, 312
 : PRIOR FILING DATE: 1994-08-23
 : PRIOR APPLICATION NUMBER: 08/210, 143
 : PRIOR FILING DATE: 1994-03-16
 : PRIOR APPLICATION NUMBER: 08/187, 757
 : PRIOR FILING DATE: 1994-01-27
 : NUMBER OF SEQ ID NOS: 78
 : SOFTWARE: PatentIn version 3.0
 : SEQ ID NO 6
 : LENGTH: 862
 : TYPE: PRT
 : ORGANISM: homo sapiens
 : US-10-079-429-6

Query Match 12.18; Score 447.5; DB 9; Length 862;
 Best Local Similarity 23.58; Pred. No. 2.7e-26;
 Matches 186; Conservative 133; Mismatches 296; Indels 175; Gaps 30;

OY 7 RGGCAGGPPR-IRRLSESVYNNRIAGVIOBPSSAAVELIENSLDAGASSVAVKDG 65
 DB 3 RAESSSTEPKAKIPIDKSVHJCSGVVLSSTAVKELENSLIDAGATNIDIKLKYG 62
 OY 66 LKLIQVSDGNGHINFEEDLALICERHTSKLSAYEDLOTISMGPRGALASMTYGVHTV 125
 DB 63 VDLIEVSDNGGVVEENFEGILLKHNHTSKIOEFADLQVETFGRGELSSICALSDYTI 122
 OY 126 TTITEGOLHGRV--STRDGVEMNEPKPCAAYKGTQVWENLFYNNVARKKTLO-NSND 182
 DB 123 STCHASAKVGRILMEFHNGKLIQKTPYR--RPRGTIVSVQOLFSTLPVRHKEFORNIKE 180
 OY 183 YPKIVDTISRVAVHINVTFSCKR--HGANRADVHASTSRLDAIRSVYGASVVDLI 239

DB 181 YAKMVOVLHAYCIIISAGINVSCTNOLQGGKROPVYCTGSGPSIKENIGSVFGQKLOSLI 240
 OY 240 -----EIKVSEDAADSIFFKADGYISNAN-VAAKIT--MLFINDRLVDC 283
 DB 241 PVOVLPPSDVCEEGGLSCDALHNLIFYISGFIISCTHVGGRSSDRPOFFINRBCDPA 300
 OY 284 ALKRAIEFYVSATLPQASKPIYMSIHLPESEVDNINHTKREVSILNDRIT----- 336
 DB 301 KXCRVLNVEYTH-MYNRHQPFFVVLINISVDECYDINVPDKKQI-LLOBEKILLAVLKTS 358
 OY 337 -----ETIRNAIEEKLAMSN 351
 DB 359 LIGMEDSDVKNLNVGQPLLDVEGNLIKMAADLEKPMVEKODQSPILRTGSEKKDVGIS 418
 OY 352 TTRITQOALNLSGIAQNPQDKYSE-----ASMSGTYSQK--IPVSOM 395
 DB 419 RLR---EAFSLRHYTEKNPSPKTPERRSPLOGRKGLMSSSTGALSDKGLVLPQKEA 474
 OY 396 VRTD--PRNPSGRLMT---YMHGSSNLEKKFDLVSVRNVVSRNQNQDAGLSR----- 446
 DB 475 VSSHGSPDPTIRAEVDESGHSTSVDSGFSIDPTGSHCSSEYAASSPGDRGQEHVD 534
 OY 447 -HELLVEIDSEFHPGLDIYKNCITYVGLADEAFALIOHNTRLYLNVVNISSKELMYQ--- 502
 DB 535 SOEKAPETDSDFS--VDCHSNQEDTGC--KFRVLPPPTNATPNTKRFKKEILLSSSD 589
 OY 503 --QALCRFENFALDSEPARPLQELLY-----MALKDELMSEKDEDE-----K 544
 DB 590 ICQKLVNTQDMSASQVDAVVKINKRVPLDPSMSIARIKQLHHEAOQSEGEQVYRRFR 649
 OY 545 LEIA-----EVNTEILK---ENAMINNEY---FSIHDOQKRLRLPVLDQYPPD 589
 DB 650 AKICGENQAADELRKELSKTMFAEMILIGFII-----TKLMEDFIYDQKHTD 704
 OY 590 ---MDRLPEVYALAGNDVYTWDEKECFRYVASAVGNFYALHPPLLPNSGNGIHLKYKN 645
 DB 705 EKYNEFMLOQHTVLOGQ-----RLIAPQTLNLTAVNEAVLI-----ENLEIFRKN 749
 OY 646 R-DSMADEHA 654
 DB 750 GFDFVIDENA 759

RESULT 13

US-09-749-601A-11
 : Sequence 11, Application US/09749601A
 : Patent No. US20020128460A1
 : GENERAL INFORMATION:
 : APPLICANT: Nicolaides, Nicholas
 : APPLICANT: Grasso, Luigi
 : APPLICANT: Saas, Philip
 : APPLICANT: Kinzler, Kenneth
 : APPLICANT: Vogelstein, Bert
 : TITLE OF INVENTION: A method for generating hypermutable
 : FILE REFERENCE: 01107.00069
 : CURRENT APPLICATION NUMBER: US/09/749, 601A
 : PRIOR FILING DATE: 2000-12-28
 : PRIOR APPLICATION NUMBER: 60/183, 333
 : PRIOR FILING DATE: 2000-02-18
 : NUMBER OF SEQ ID NOS: 14
 : SOFTWARE: FastSeq for Windows Version 4.0
 : SEQ ID NO 11
 : LENGTH: 862
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 : US-09-749-601A-11

Query Match 12.18; Score 447.5; DB 10; Length 862;
 Best Local Similarity 23.58; Pred. No. 2.7e-26;
 Matches 186; Conservative 133; Mismatches 296; Indels 175; Gaps 30;

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 15:45:41 ; Search time 40 Seconds

(without alignments)
2411.836 Million cell updates/sec

Title: US-09-954-950-2

Perfect score: 724
Sequence: 1 MDESPRGCGCAGPEPRIR.....DGFVVASLEKLYIFERC 724

Scoring table: OLIGO
Gapop 60.0 , Gapept 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_101002:*

- 1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	724	100.0	724	23	AAE22978
2	19	2.6	737	22	AAE08710
3	19	2.6	737	23	AAE22879
4	14	1.9	752	18	AAW09036
5	14	1.9	756	16	AAE75785
6	14	1.9	756	14	AAE76071
7	14	1.9	756	16	AAE79008
8	14	1.9	756	22	AAE85854
9	14	1.9	756	22	AAE63956
10	14	1.9	756	23	AAE24359

11	14	1.9	756	23	AAE98778	Human Mutl homolog
12	14	1.9	775	22	AAE75536	Human colon cancer
13	13	1.8	133	22	AAE85855	Human PMS2-134 pro
14	13	1.8	133	22	AAE63957	Amino acid sequenc
15	13	1.8	442	20	AAE37635	Protein involved i
16	13	1.8	580	20	AAE35452	Chlamydia pneumoni
17	13	1.8	900	22	AAE85851	Human PMS2 protein
18	13	1.8	931	16	AAE79009	Human DNA repair p
19	13	1.8	932	22	AAE85852	Human PMS1 protein
20	13	1.8	932	22	AAE63953	Amino acid sequenc
21	13	1.8	932	22	AAE63954	Amino acid sequenc
22	13	1.8	932	23	AAE24357	Human mismatch rep
23	13	1.8	932	23	AAE24357	Human mismatch rep
24	9	1.2	54	19	AAE44347	Human post meiotic
25	9	1.2	76	23	AAE06237	Human ORFX protein
26	9	1.2	363	22	AAE48199	Staphylococcus epi
27	9	1.2	363	22	AAE48199	Mature caastor eny
28	9	1.2	391	13	AAE20011	Soybean seed stea
29	9	1.2	391	16	AAE82024	Stearyl-ACP-desat
30	9	1.2	391	19	AAE37939	Amino acid sequenc
31	9	1.2	396	12	AAE14188	C. tinctorius delt
32	9	1.2	396	13	AAE22048	Carthamus tinctori
33	9	1.2	396	15	AAE54033	Sequence encoding
34	9	1.2	396	15	AAE44349	Carthamus tinctori
35	9	1.2	396	19	AAE44350	Ricinus communis d
36	9	1.2	629	22	AAE35368	Haemophilus influe
37	9	1.2	629	22	AAE35368	Pseudomonas aerugi
38	9	1.2	664	22	AAE65767	Drosophila melanog
39	9	1.2	669	22	AAE33766	Staphylococcus aur
40	9	1.2	669	22	AAE36741	Staphylococcus aur
41	9	1.2	769	22	AAE85849	Yeast MLH1 protein
42	9	1.2	769	22	AAE63951	Amino acid sequenc
43	9	1.2	859	22	AAE85850	Mouse PMS2 protein
44	9	1.2	859	22	AAE63952	Amino acid sequenc
45	9	1.2	859	23	AAE24355	Mouse mismatch rep

ALIGNMENTS

RESULT 1	AAE22978	
ID	AAE22978	standard; Protein; 724 AA.
XX		
AC	AAE22978;	
XX		
DT	21-AUG-2002	(first entry)
XX		
DE	Rice MLH1 protein.	
XX		
KW	Rice; MLH1: gene mutation; mismatch repair system; transposon tagging;	
KW	tissue-specific promoter; herbicidal safener; male sterility; detection;	
KW	pollen formation; base pair mismatch.	
XX		
OS	Oryza sativa.	
XX		
FH	Key	Location/Qualifiers
FT	Region	103..119
FT		/note="mutl/PMS1 signature sequence"
XX		
PN	WO200224890-A2.	
XX		
PD	28-MAR-2002.	
XX		
PF	18-SEP-2001; 2001MC-US29088.	
XX		
PR	18-SEP-2000; 2000US-233124P.	
XX		
PA	(PION-) PIONEER HI-BRED INT INC.	
XX		
PI	Mahajan PB;	
XX		
DR	WPI; 2002-416283/44.	

DR N-PSDB; AAD36728.

PT Novel rice MLH1 ortholog nucleic acid molecule for increasing
PT efficiency of targeted gene mutation or homologous recombination in a
PT plant and for generating plants with reversible male sterility

PS Claim 7; Fig 2; 90pp; English.

xx The invention relates to isolated rice MLH1 orthologue nucleic acids. The
cc nucleic acid is useful for increasing the efficiency of targeted gene
cc mutation or homologous recombination in a plant, by transforming a plant
cc with expression cassette comprising the nucleic acid linked to a chemical
cc inducible promoter, transforming the plant with nucleic acid comprising
cc a sequence having a desired mutation or a sequence to be homologously
cc recombined, where the transformation occurs in the presence of chemical
cc compound capable of inducing the promoter and the plant's cellular
cc mismatch repair system is inhibited and selecting the transformed plants
cc that contained the mutation or homologously recombined nucleotide
cc sequence. The plant cellular mismatch repair system is inhibited through
cc the use of transposon tagging of an MLH1 gene, sense- and antisense-
cc suppression of an MLH1 gene, antibody binding to an MLH1 polypeptide or
cc its variant, and targeted mutagenesis of specific amino acid residues
cc encoded by an MLH1 gene. The nucleic acid is also useful for producing
cc a reversible male sterility in a plant, by transforming a plant with an
cc expression cassette comprising a lexa DNA binding site embedded in a
cc tissue-specific promoter that drives expression in the plant operably
cc linked to the nucleic acid when expressed disrupts pollen formation or
cc function through inhibition of the plant's cellular mismatch repair
cc system, transforming the plant with a second expression cassette
cc comprising a nucleotide sequence encoding a lexa repressor protein
cc operably linked to a chemically-inducible promoter that drives expression
cc in the plant, and exposing the plant to a compound capable of inducing
cc the chemical-inducible promoter, to induce expression of lexa repressor
cc protein. The tissue-specific promoter is an anther-specific promoter
cc and the chemical-inducible promoter is a herbicidal safener. The
cc polypeptide encoded by the nucleic acid is useful for detecting,
cc locating, or removing a base pair mismatch (SNP). The present sequence
cc is rice MLH1 protein.

xx Sequence 724 AA:

Query Match 100.0%; Score 724; DB 23; Length 724;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 724; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSPSRGGCGAGGPRIRIRLESYVNRITAGVIOBPSSAVKELIENSLDAGASSVYA 60
DB 1 MDSPSRGGCGAGGPRIRIRLESYVNRITAGVIOBPSSAVKELIENSLDAGASSVYA 60
QY 61 VKDGGKLIQVSDGHIREFEDIALICERHTTSKLSAYEDLOTIKSMGRGELASMTYV 120
DB 61 VKDGGKLIQVSDGHIREFEDIALICERHTTSKLSAYEDLOTIKSMGRGELASMTYV 120
QY 121 GHVTTVTITBGLHGYRSTYRGVMEVNEKRPCAAVGTGVYVNEFLYNNVARKTIOONS 180
DB 121 GHVTTVTITBGLHGYRSTYRGVMEVNEKRPCAAVGTGVYVNEFLYNNVARKTIOONS 180
QY 121 GHVTTVTITBGLHGYRSTYRGVMEVNEKRPCAAVGTGVYVNEFLYNNVARKTIOONS 180
DB 121 GHVTTVTITBGLHGYRSTYRGVMEVNEKRPCAAVGTGVYVNEFLYNNVARKTIOONS 180
QY 181 DDVPKIVDFSRFAVHNINVTFCRKHGANRADVHSASTSSRDAIRSYVYASVVDLIE 240
DB 181 DDVPKIVDFSRFAVHNINVTFCRKHGANRADVHSASTSSRDAIRSYVYASVVDLIE 240
QY 241 IKVSYEDADADSIKMDGYISNANYVAKITMILFINDRLVDTALRAIEYVSATLPOA 300
DB 241 IKVSYEDADADSIKMDGYISNANYVAKITMILFINDRLVDTALRAIEYVSATLPOA 300
QY 301 SKPFIYMSIHLEBEHVDVNIHPKKEVSLNORRIETIRNAIEEKLMSNTTRIFOTQA 360
DB 301 SKPFIYMSIHLEBEHVDVNIHPKKEVSLNORRIETIRNAIEEKLMSNTTRIFOTQA 360
QY 361 LNTSGIAQNPQKDKVSEASMGSGTSOKIPVSOVMTDPRNSGRLHTYWHGOSSNLEK 420
DB 361 LNTSGIAQNPQKDKVSEASMGSGTSOKIPVSOVMTDPRNSGRLHTYWHGOSSNLEK 420

QY 421 KFDLVSVNRVVRNRKNDGADLSRRHELLVEIDSSFPGLDVIYVNCYVGLADAFAL 480
DB 421 KFDLVSVNRVVRNRKNDGADLSRRHELLVEIDSSFPGLDVIYVNCYVGLADAFAL 480
QY 481 IOHNTRLVYVNVVNIISKELMYOQALCRFNFNAIQUSEPAPLOELLVMAIKDELMSDEK 540
DB 481 IOHNTRLVYVNVVNIISKELMYOQALCRFNFNAIQUSEPAPLOELLVMAIKDELMSDEK 540
QY 541 DDEKLEIAEYNTLEIKNAEMINEYFSIHIDOGKTLRPLVDOTYTPMDRLPEYVAL 600
DB 541 DDEKLEIAEYNTLEIKNAEMINEYFSIHIDOGKTLRPLVDOTYTPMDRLPEYVAL 600
QY 601 GNDVTDDEKECFRTVASAVGNFYALHPITLPNPSGNGIHLVYKNRDSMADEHAENDLIS 660
DB 601 GNDVTDDEKECFRTVASAVGNFYALHPITLPNPSGNGIHLVYKNRDSMADEHAENDLIS 660
QY 661 DENDVDELLAEAAVANAQREWTIOHVLFPSSMRFLFKPPSMATDGTFFVOVASIEKLYKI 720
DB 661 DENDVDELLAEAAVANAQREWTIOHVLFPSSMRFLFKPPSMATDGTFFVOVASIEKLYKI 720
QY 721 FERC 724
DB 721 FERC 724

RESULT 2
AAE08710
ID AAE08710 standard; Protein; 737 AA.

AC AAE08710;

DT 15-NOV-2001 (first entry)

DE Arabidopsis thaliana PMS2 protein homologue MLH.

KM Hypermutable plant; dominant negative allele; mismatch repair gene;

KW MMR; cell line generation; PMS2; ACDMLH.

OS Arabidopsis thaliana.

PN M0200161012-AL.

PD 23-AUG-2001.

PF 28-DEC-2000; 2000WO-US35397.

PR 18-FEB-2000; 2000US-0183333.

PA (NICO/) NICOLAIDES N C.

PA (GRAS/) GRASSO L.

PA (SASS/) SASS P M.

PA (KINZ/) KINZLER K.

PA (VOGE/) VOGELSTEIN B.

PI Nicolaides NC, Grasso L, Sass PM, Kinzler K, Vogelstein B;

DR WPI; 2001-529913/58.

PS Example 1; Page 57-59; 72pp; English.

CC The invention relates to a method for generating hypermutable plant,

CC especially crop plants with new output traits, comprises introducing

CC polynucleotide comprising dominant negative allele of mismatch repair

CC gene into plant cell

CC commercial use. The present sequence is Arabidopsis thaliana (At)

CC MLH protein. This sequence is a homologue of MMR protein, PMS2.
XX
SQ Sequence 737 AA;

Query Match 2.6%; Score 19; DB 22; Length 737;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 RLEESVNNRIAGEVIQRP 38
DB 31 RLEESVNNRIAGEVIQRP 49

RESULT 3

ID AAE22979 standard; Protein; 737 AA.

XX AAE22979;

DT 21-AUG-2002 (first entry)

XX Arabidopsis thaliana MLH1 protein.

XX MLH1: gene mutation; mismatch repair system; transposon tagging;
KM tissue-specific promoter; herbicidal safener; male sterility; detection;
KM pollen formation; base pair mismatch.

XX Arabidopsis thaliana.

PN WO200224890-A2.

PD 28-MAR-2002.

PF 18-SEP-2001; 2001WO-US29088.

PR 18-SEP-2000; 2000US-233124P.

PA (PION-) PIONEER HI-BRED INT INC.

PI Mahajan PB;

DR WPI: 2002-416283/44.

DR N-PSDB; AAD36729.

XX Novel rice MLH1 ortholog nucleic acid molecule for increasing
PT efficiency of targeted gene mutation or homologous recombination in a
PT plant and for generating plants with reversible male sterility

XX Disclosure; Page 89-90; 90pp; English.

XX The invention relates to isolated rice MLH1 orthologue nucleic acids. The
CC nucleic acid is useful for increasing the efficiency of targeted gene
CC mutation or homologous recombination in a plant, by transforming a plant
CC with expression cassette comprising the nucleic acid linked to a chemical
CC inducible promoter, transforming the plant with nucleic acid comprising
CC a sequence having a desired mutation or a sequence to be homologously
CC recombined, where the transformation occurs in the presence of chemical
CC compound capable of inducing the promoter and the plant's cellular
CC mismatch repair system is inhibited and selecting the transformed plants
CC that contained the mutation or homologously recombined nucleotide
CC sequence. The plant cellular mismatch repair system is inhibited through
CC the use of transposon tagging of an MLH1 gene, sense- and antisense-
CC suppression of an MLH1 gene, antibody binding to an MLH1 polypeptide or
CC its variant, and targeted mutagenesis of specific amino acid residues
CC encoded by an MLH1 gene. The nucleic acid is also useful for producing
CC reversible male sterility in a plant, by transforming a plant with an
CC expression cassette comprising a lex A DNA binding site embedded in a
CC tissue-specific promoter that drives expression in the plant operably
CC linked to the nucleic acid when expressed disrupts pollen formation or
CC function through inhibition of the plant's cellular mismatch repair
CC system, transforming the plant with a second expression cassette
CC comprising a nucleotide sequence encoding a lex A repressor protein
CC operably linked to a chemically-inducible promoter that drives expression

CC in the plant, and exposing the plant to a compound capable of inducing
CC the chemical-inducible promoter, to induce expression of lex A repressor
CC protein. The tissue-specific promoter is an anther-specific promoter
CC and the chemical-inducible promoter is a herbicidal safener. The
CC polypeptide encoded by the nucleic acid is useful for detecting,
CC localizing, or removing a base pair mismatch (SNP). The present sequence
CC is Arabidopsis thaliana MLH1 protein.

SQ Sequence 737 AA;

Query Match 2.6%; Score 19; DB 23; Length 737;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 RLEESVNNRIAGEVIQRP 38
DB 31 RLEESVNNRIAGEVIQRP 49

RESULT 4

AAW09036 standard; protein; 752 AA.

XX AAW09036;

DT 01-OCT-1997 (first entry)

XX Mismatch repair protein, hPMS2.

XX Human; mismatch repair protein; hMSH2, hMLH1; hPMS2; antibody; Ab;
KM proliferation; anti-cancer therapy; neoplastic tissue; DNA replication;
KM DNA repair; mutation; sporadic colorectal carcinoma; cancer; HNPCC;
XX inherited; non-polyposis colorectal carcinoma.

OS Homo sapiens.

PN WO9641192-A1.

PD 19-DEC-1996.

PF 07-JUN-1996; 96WO-US08957.

PR 07-JUN-1995; 95US-0480351.

PA (UYJO) UNIV JOHNS HOPKINS.

PI Burrell M, Hill DE, Kinzler KW, Vogelstein B;

DR WPI: 1997-077287/07.

XX Discriminating proliferating from non-proliferating cells in tissue
PT - using antibodies specifically immuno-reactive with mismatch repair
PT protein, esp. human MSH2

XX Disclosure; Page 23-25; 37pp; English.

XX The sequences given in AAW09034-36 represent the human mismatch repair
CC proteins, hMSH2, hMLH1 and hPMS2. In the method of the invention, these
CC proteins were identified by reaction with an antibody (Ab) specific for
CC them, therefor discriminating proliferating from non-proliferating cells
CC The method may be used for monitoring the effectiveness of anti-cancer
CC therapy in neoplastic tissue, by comparing the amount of Ab-Ag complexes
CC in the sample with an amount determined at an earlier time, in which a
CC reduction in the amount indicates an effective therapy. The Ab are
CC especially specifically immunoreactive with the MSH2 mismatch repair
CC gene, which is 1 of at least 4 genes encoding proteins involved in the
CC repair of mismatched nucleotides following DNA replication or repair.
CC Mutations in the MSH2 gene contribute to the development of sporadic
CC colorectal carcinoma, while germline MSH2 mutations are responsible for
CC approx. 50% of inherited, non-polyposis colorectal carcinoma (HNPCC).
CC Since MSH2 is ubiquitously expressed, development of other cancers are
CC also susceptible to alterations in MSH2.

SO Sequence 752 AA;

Query Match 1.9%; Score 14; DB 16; Length 752;
Best Local Similarity 100.0%; Pred. No. 3.1e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 VVNR1AGEV1ORP 38
DB 15 VVNR1AGEV1ORP 28

RESULT 5
ID AAR75785 standard; Protein: 756 AA.
XX AAR75785;
AC AAR75785;
DT 04-MAR-1996 (first entry)
XX Human wild type MLH1, a mult. homologue.
XX hMLH1; wild type; mult. homologue; cancer diagnosis; mismatch repair;
KM tumour; susceptibility; mutation detection.
XX Homo sapiens.
OS
XX W09516793-A1.
PN 22-JUN-1995.
PD 16-DEC-1994; 94MO-US14746.
PF 09-DEC-1994; 94US-035902.
PR 17-DEC-1993; 93US-0168877.
PR 08-MAR-1994; 94US-0209521.
XX (DAND) DANA FARBER CANCER INST INC.
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX Baker SM, Bollag RJ, Bronner CE, Kolodner RD, Lisakay RM;
DR WPI: 1995-231583/30.
N-PSDB; AAO90814.
XX
XX Determ. of a mutation in a mult. homologue or gene prod. in a tissue
PT used to diagnose cancer susceptibility, and to identify and
PT classify a DNA mismatch-repair-defective tumour
PS Claim 33; Fig 3; 168pp; English.
XX
XX AAO90814 encodes AAR75785 the wild type hMLH1, a mult. homologue. A
CC mutation in a hMLH1 or hPMS1 nucleic acid isolated from a subject,
CC can be detected by comparing it with an analogous segment of the
CC above wild type allele. This method can be used to diagnose cancer
CC susceptibility, or to identify and classify a DNA mismatch-repair
CC defective tumour.
XX
XX Sequence 756 AA;

Query Match 1.9%; Score 14; DB 16; Length 756;
Best Local Similarity 100.0%; Pred. No. 3.1e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 VVNR1AGEV1ORP 38
DB 15 VVNR1AGEV1ORP 28

RESULT 6
ID AAR76071 standard; Protein: 756 AA.
XX AAR76071;
AC AAR76071;

XX 15-JAN-1996 (first entry)
XX Human mismatch repair pathway protein, Mh1.
DE Mismatch repair; MSH2; primer; identification; defect; alteration;
XX Cancer; tumour; vaccine.
XX Homo sapiens.
OS
XX W09514085-A2.
PN 26-MAY-1995.
PD 17-NOV-1994; 94MO-US13385.
PF 13-JUN-1994; 94US-0259310.
PR 17-NOV-1993; 93US-0154792.
PR 07-DEC-1993; 93US-0163449.
XX (UYVE-) UNIV VERMONT & STATE AGRIC COLLEGE.
PA (DAND) DANA FARBER CANCER INST.
XX Fishel R, Kolodner RD, Reenan RAG;
PI WPI: 1995-200377/26.
DR N-PSDB; AAO94015.
XX
XX Determining alteration in human mismatch repair pathways - used in
PT the diagnosis, prognosis and therapy of cancers and in screening
PT assays
XX
XX Example 13; Page 239-242; 256pp; English.
XX
XX AAR76071 is the human mismatch repair pathway protein Mh1. Defects or
CC alterations in a human mismatch repair gene results in the accumulation
CC of unstable repeated DNA sequences, a feature of a number of different
CC cancers. The identification of a defect in the mismatch repair pathway
CC can be diagnostic of a predisposition to cancer and prognostic for a
CC particular mammalian cancer e.g colorectal, ovarian, endometrial
CC (uterine), renal, bladder, skin, rectal and bowel. The nucleotide
CC sequences and polypeptides of the hMSH2 gene may also be used for
CC therapy and in vaccines.
XX
XX Sequence 756 AA;

Query Match 1.9%; Score 14; DB 16; Length 756;
Best Local Similarity 100.0%; Pred. No. 3.1e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 VVNR1AGEV1ORP 38
DB 15 VVNR1AGEV1ORP 28

RESULT 7
ID AAR79008 standard; Protein: 756 AA.
XX AAR79008;
AC AAR79008;
DT 26-MAR-1996 (first entry)
XX Human DNA repair protein hMLH1.
XX DNA repair protein; hMLH1; hMLH2; hMLH3; therapy; cancer; vectors;
KM DNA synthesis; diagnosis; disease; mult4.
XX
XX Homo sapiens.
OS
XX W09520678-A1.
PN 03-AUG-1995.
PD

```
XX 25-JAN-1995; 95WO-US01035.
PF 23-AUG-1994; 94US-0294312.
XX 27-JAN-1994; 94US-0187757.
PR 16-MAR-1994; 94US-0210143.
XX (HOMA-) HUMAN GENOME SCI INC.
PA Adams MD, Fleischmann RD, Fraser CM, Fulmer RA;
PI Haseltine WA, Kirtness EF, Rosen CA, Ruben SM, Wei Y;
XX WPI: 1995-275461/36.
DR N-PSDB; AAG97525.
XX Polynucleotide(s) encoding human mutL4 homologues, hMLH1, hMLH2 and
PT hMLH3 - used for therapeutic treatment of, e.g. hereditary cancer
XX Claim 8; Figure 1; 124pp: English.
XX The polynucleotides described in AAG97525-27 encode the human
CC analogues of the prokaryotic mutL4 DNA repair gene. The polypeptides
CC they encode (AAR79008-R79010) are used for therapeutic purposes e.g.
CC in the treatment of cancer, esp. hereditary cancer. They may also
CC be used for in vitro manipulation of DNA, synthesis of DNA and the
CC manufacture of DNA vectors and in methods of diagnosing a disease or
CC a susceptibility to a disease related to a mutation in the hMLH1, -2
CC or -3 DNA repair genes.
XX
SQ Sequence 756 AA:
Query Match 1.9%; Score 14; DB 16; Length 756;
Best Local Similarity 100.0%; Pred. No. 3.1e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 25 VVNR1AAGEV10RP 38
DB 15 VVNR1AAGEV10RP 28
RESULT 8
AAB85854
ID AAB85854 standard; Protein; 756 AA.
XX AAB85854;
AC
XX 29-OCT-2001 (first entry)
DT
XX Human MLH1 protein.
DE
XX Hypermutable bacteria; mismatch repair gene; MMR gene; Muth; Muts; MutL;
KM Mut; PMS2; MLH1; MLH3; PMSR; biocatalysis; bioremediation; biochemical;
KW drug discovery; detoxification; toxin; biotransformation.
XX
OS Homo sapiens.
XX
XX WO200159092-A2.
XX
XX 16-AUG-2001.
PD
XX 12-FEB-2001; 2001WO-US04339.
PF
XX 11-FEB-2000; 2000US-0181929.
PR
XX (UYJO) UNIV JOHNS HOPKINS.
PA
XX Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW;
PI WPI: 2001-514664/56.
DR N-PSDB; AAG76367.
XX
XX Making hypermutable bacteria for biocatalysis, bioremediation and drug
PT discovery, involves introducing polynucleotide comprising dominant
```

```
PT negative allele of mismatch repair gene under regulatory sequence
PT control
XX Example 1; Page 44; 68pp: English.
PS The invention provides a method for generating a hypermutable bacteria.
XX The method involves introducing a polynucleotide having a dominant
CC negative allele of a mismatch repair (MMR) gene under the control of an
CC inducible transcription regulatory sequence, into a bacterium. The cell
CC becomes inducibly hypermutable. The method is useful to create desirable
CC output traits for commercial applications, using dominant negative
CC alleles of mismatch repair proteins. The mismatch repair gene is a Muth,
CC Muts, MutL or MutR homologue and can be selected from PMS2, MLH1, MLH3,
CC PMSR or PMSR homologue. The hypermutable bacteria is useful for the
CC production, biocatalysis, bioremediation and drug discovery. It is also
CC useful in manufacturing industry for the generation of new biochemicals
CC useful for detoxifying noxious chemicals from by-products of
CC manufacturing processes or those used as catalysts, for remediation of
CC toxins present in the environment including polychlorobenzenes, heavy
CC metals and other environmental hazards for which there is a need to
CC remove them from the environment. The hypermutable bacteria is also
CC useful for screening novel mutations in a gene or a set of genes that
CC produce variant siblings that exhibit a new output trait not found in
CC wild type cells. The bacteria are also useful for producing increased
CC quantity or quality of protein or non-protein therapeutic molecule e.g.
CC penicillin G, Erythromycin and Clavulanic acid, by biotransformation.
CC Dominant negative alleles of the MMR gene are useful for producing higher
CC quantities of recombinant polypeptides. The present sequence represents
CC a human MLH1 protein.
XX
SQ Sequence 756 AA:
Query Match 1.9%; Score 14; DB 22; Length 756;
Best Local Similarity 100.0%; Pred. No. 3.1e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 25 VVNR1AAGEV10RP 38
DB 15 VVNR1AAGEV10RP 28
RESULT 9
AAG63956
ID AAG63956 standard; Protein; 756 AA.
XX AAG63956;
AC
XX 29-OCT-2001 (first entry)
DT
XX Amino acid sequence of human mismatch repair protein MLH1.
DE
XX MLH1; mismatch repair gene; MMR gene; hypermutable yeast.
KM
XX
XX Homo sapiens.
XX
XX WO200162945-A1.
XX
XX 30-AUG-2001.
PD
XX 21-FEB-2001; 2001WO-US05447.
PF
XX 23-FEB-2000; 2000US-0184336.
PR
XX (UYJO) UNIV JOHNS HOPKINS.
PA (NICO/) NICOLAIDES N C.
XX (SASS/) SASS P M.
PA (GRAS/) GRASSO L.
XX (VOGE/) VOGELSTEIN B.
PA (KINZ/) KINZLER K W.
XX
XX Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW;
PI WPI: 2001-52820/57.
DR
```


CC composition is useful for obtaining a mammalian cell that is resistant
CC to a selected microbe. (M1) involves growing a culture of mammalian
CC cells with a dominant negative allele of a mismatch repair gene,
CC exposing the cells to the selected microbe, and selecting the mammalian
CC cell resistant to the selected microbe. The hypermutable cell is
CC resistant to gram-negative and gram-positive microbe, protozoan,
CC bacteria or fungi. The microbial resistance is selected by isolating
CC and testing conditioned medium from the hypermutable cell. The
CC composition is also useful for obtaining a cell comprising a mutation
CC in a gene encoding an antimicrobial activity. (M1) involves growing a
CC culture of mammalian cells with the gene encoding the antimicrobial
CC activity, and a dominant negative allele of a mismatch repair gene,
CC selecting a cell comprising the antimicrobial activity, and determining
CC whether the gene comprises a mutation. The cell is examined to
CC determine whether the gene comprises a mutation by analysing a
CC nucleotide sequence of the gene or mRNA transcribed from the gene, a
CC protein encoded by gene or its phenotype. The composition
CC is useful for identifying antimicrobial agents, microbe-specific toxic
CC molecules, and for producing new phenotypes of the cell. (M1) is useful
CC for creating genetically altered antimicrobial molecules, and also for
CC creating cell lines that manufacture antimicrobial molecules for use in
CC large scale production of antimicrobial agents for clinical studies.
CC (M1) is also useful in cell lines that express known antimicrobial
CC agents to enhance the biochemical activity of the antimicrobial agent.
CC The present sequence represents human MLH1.

SQ Sequence 756 AA;

Query Match

1.9%; Score 14; DB 23; Length 756;

Best Local Similarity 100.0%; Pred. No. 3.1e-05;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 VVNRIAGGEVQRP 38
|||||
Db 15 VVNRIAGGEVQRP 28

RESULT 12

AAG75536

ID AAG75536 standard; Protein; 775 AA.

AC AAG75536;

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO:6300.

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;

KM colorectal carcinoma; chromosome 3.

OS Homo sapiens.

PN WO200122920-A2.

PD 05-APR-2001.

PF 28-SEP-2000; 2000WO-US26524.

PR 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Barash SC, Birse CE, Rosen CA;

DR WPI: 2001-235357/24.

DR N-PSDB; AAH34941.

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
PS Claim 11: Page 7756-7760; 9803pp; English.

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated P,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

SQ Sequence 775 AA;

Query Match

1.9%; Score 14; DB 22; Length 775;

Best Local Similarity 100.0%; Pred. No. 3.2e-05;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 VVNRIAGGEVQRP 38
|||||
Db 34 VVNRIAGGEVQRP 47

RESULT 13

AAB85855

ID AAB85855 standard; Protein; 133 AA.

AC AAB85855;

DT 29-OCT-2001 (first entry)

DE Human PMS2-134 protein.

KW Hypermutable bacteria; mismatch repair gene; MMR gene; Muth; Muts; Mutt;

KW Muty; PMS2; MLH1; MLH3; PMSR; biocatalysis; bioremediation; biochemical;

KM drug discovery; detoxification; toxin; biotransformation; PMS2-134.

OS Homo sapiens.

PN WO200159092-A2.

PD 16-AUG-2001.

PF 12-FEB-2001; 2001WO-US04339.

PR 11-FEB-2000; 2000US-0181929.

PA (UYJO) UNIV JOHNS HOPKINS.

PI Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW;

DR WPI: 2001-514664/56.

DR N-PSDB; AAH76368.

PT Making hypermutable bacteria for biocatalysis, bioremediation and drug
PT discovery, involves introducing polynucleotide comprising dominant
PT negative allele of mismatch repair gene under regulatory sequence
PT control -
PS Example 1: Page 45; 68pp; English.

CC The invention provides a method for generating a hypermutable bacteria.
CC The method involves introducing a polynucleotide having a dominant
CC negative allele of a mismatch repair (MMR) gene under the control of an
CC inducible transcription regulatory sequence, into a bacterium. The cell

```

AA      sequence      133 AA;
SQ
Query Match      1.8%; Score 13; DB 22; Length 133;
Best Local Similarity 100.0%; Pred. No. 6.4e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

```

RESULT 14
AAG63957
ID AAG63957 standard; Protein; 133 AA

XX 29-OCT-2001 (first entry)
DT

XX DE Amino acid sequence of human mismatch repair protein PMS2-134

XX mismatch repair gene; MMR gene; hypermutable yeast
KW PMS2-134; mismatch repair gene; MMR gene; hypermutable yeast

XX OS Homo sapiens

XX
PN WO200162945-A1

XX 30-AUG-2001.
PD

XX
PF 21-FEB-2001; 2001WO-US05447.

XX 23-FEB-2000; 2000US-0184336
PR
XX

AA
PA
PA
(UYJO) UNIV JOHNS HOPKINS
(NICO/) NICOLAIDES N C.

PA	(GRASS/)	GRASSO L.
PA	(SASS/)	SASS P M.
PA	(NICO/)	NICO P M.

PA (KIN%) KINZLER K W.
PA (VOGE/) VOGELSTEIN B
PA (GAND/) GANDER Z

XX Nicolaides NC, Sassi PM, Grasso L, Vogelstein B, Kinzler KW

XX
DR WPI: 2001-522820/57

DR N-PSDB; AAH75045
XX

PT Making hypermutable yeast that exhibit novel selected output traits for commercial applications, comprises introducing polynucleotide

PT containing dominant negative allele of mismatch repair gene
XX
PS Disclosure; Page 45; 60pp: English.

XX The present sequence represents human PMS2-134. PMS2-134 is a mismatch
CC

[illegible]

RESULT 15
AAV37635
ID AAV37635 standard; Protein; 442 AA

XX
DT 07-OCT-1999 (first entry)

Protein involved in intermediate metabolism of nucleic acids

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KM Trachoma; conjunctivitis; genital disease; periorbitis;

KW	paratrachoma;	inclusion conjunctivitis; genital chlamydiosis
KW	nongonococcal urethritis; epididymitis; cervicitis; salpingitis	
KW	gonorrhea; gonococci; lymphogranulomatosis.	

Chlamydia trachomatis
bacterial conjunctivitis, pneumoniae

03
XX
PN
W099928475-A2

10-JUN-1999

XX 27-NOV-1998; 98WO-IB01939

XX
PR 04-NOV-1998; 98US-0107077

PR	28-NOV-1997;	97FR-0015041
PR	17-DEC-1997;	97FR-0016034

XX
PA (GEST) GENSET.

XX
PI Griffais R;

XX
DR WPI; 1999-371125/31

XX
PT Genome sequence of *Chlamydia trachomatis*

XX
PS Disclosure; Page 1273-1274; 1755pp; English

XX
CC
AAV36754-V37949 are encoded by open reading frames (ORFs) of the genome (AAV36754-V37949). The polyomaviruses can be used as

CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC (see also below) to control growth of the microorganism. Chlamydia
CC of Chlamydia trachomatis (see 4404423). The polypeptide and amino acid
CC sequences of Chlamydia trachomatis. Antisense and ribozyme sequences

trachomatis is responsible for a large number of diseases, e.g. eye diseases such as conventional trachoma, nonendemic trachoma, CC

SUCH AS
DISEASES

CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC perihepatitis, bartolinitis; pneumonia in breast feeding infants;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.

XX
 SQ Sequence 442 AA;

Query Match 1.8%; Score 13; DB 20; Length 442;
 Best Local Similarity 100.0%; Pred. No. 0.00021;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 VKELIENSIDAGA 54
 |||||H|||||||
 DB 32 VKELIENSIDAGA 44

Search completed: March 27, 2003, 15:51:23
 Job Time : 42 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 15:42:51 ; Search time 26 Seconds

(without alignments)
2676.974 Million cell updates/sec

Title: US-09-954-950-2

Perfect score: 3709

Sequence: 1 MDEPSPRGCGAGCPPPIRR.....DGTGVVASLEKLYKIERC 724

Scoring table: BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	2505	67.5	737	2	MLH1 protein [limo
2	2505	67.5	737	2	DNA mismatch repair
3	1452	39.1	756	2	DNA mismatch repair
4	1097	29.6	769	2	DNA mismatch repair
5	1040.5	28.1	684	2	probable DNA mismatch repair
6	845	22.8	779	2	hypothetical protein
7	597	16.1	669	2	DNA mismatch repair
8	584.5	15.8	622	2	DNA mismatch repair
9	565	15.2	649	2	DNA mismatch repair
10	565	15.2	649	2	DNA mismatch repair
11	560	15.1	649	2	DNA mismatch repair
12	555	15.0	695	2	DNA mismatch repair
13	547.5	14.8	629	2	DNA mismatch repair
14	538	14.5	658	2	DNA mismatch repair
15	532	14.3	603	2	DNA mismatch repair
16	529	14.3	658	2	DNA mismatch repair
17	523	14.1	601	2	DNA mismatch repair
18	509.5	13.7	627	2	DNA mismatch repair
19	505	13.6	637	2	DNA mismatch repair
20	486.5	13.1	618	2	DNA mismatch repair
21	482.5	13.0	618	2	DNA mismatch repair
22	471.5	12.7	633	2	DNA mismatch repair
23	465.5	12.6	610	2	DNA mismatch repair
24	461	12.4	609	2	DNA mismatch repair
25	461	12.4	615	2	DNA mismatch repair
26	461	12.4	615	2	DNA mismatch repair
27	461	12.4	628	2	DNA mismatch repair
28	461	12.4	653	2	DNA mismatch repair
29	459	12.4	613	2	DNA mismatch repair

30	459	12.4	635	2	AC0046	DNA mismatch repair
31	456.5	12.3	610	2	A97870	DNA mismatch repair
32	456	12.3	659	2	E84176	DNA mismatch repair
33	448	12.1	623	2	A13637	DNA mismatch repair
34	447.5	12.1	862	2	S47598	mult. protein homolog
35	446.5	12.0	794	2	T37989	DNA mismatch repair
36	437.5	11.8	516	2	H72427	DNA mismatch repair
37	433.5	11.7	584	2	D84996	DNA mismatch repair
38	433.5	11.7	595	2	F71650	DNA mismatch repair
39	426.5	11.5	619	2	E82765	DNA mismatch repair
40	425.5	11.5	580	2	A73032	DNA mismatch repair
41	425.5	11.5	580	2	B86592	DNA mismatch repair
42	415.5	11.2	576	2	A71497	probable DNA mismatch repair
43	415.5	11.2	637	2	D87335	probable DNA mismatch repair
44	414	11.2	648	2	A71342	probable DNA mismatch repair
45	409.5	11.0	904	2	S53896	DNA mismatch repair

ALIGNMENTS

RESULT 1

F85092

MLH1 protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Mar-2001

C:Accession: F85092

R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: AB5001; MUID:20083488; PMID:10617198

A:Accession: F85092

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-737 <STO>

A:Cross-References: GB:NC_001268; NID:g7267557; PIDN:CAB78038.1; GSPDB:GN00140

A:Gene: AT4G09140

A:Map position: 4

C:Superfamily: mismatch repair protein MLH1

Query Match	Score	DB 2;	Length	737;
Best local similarity	67.5%	Pred. No. 5.5e-146;		
Matches	482;	Conservative 111;	Mismatches 129;	Indels 4;
Gaps	3;			
QY	2	DEPSPRGCGAGCPPPIRRLEESVNRIRAGETIOPSSAVKELIENSIDAGASSVAV 61		
DB	13	EEESPATTIVPRPKIQRLSESVNRIRAGETIOPSSAVKELIENSIDAGASSVAV 72		
QY	62	KDGLKLIQVSDGHRGIRFEDLILCERHTTSKLASAVEDLOTIKSGFGEALASMTYVG 121		
DB	73	KDGLKLIQVSDGHRGIRFEDLILCERHTTSKLIFEDLIFSLSGFGEALASMTYVA 132		
QY	122	HVTVTTFEGQLHGIVSVYRDGVMEHPKCAAVKGTQVWENLFYNNVARKKTLQNSND 181		
DB	133	HVTVTTFEGQLHGIVSVYRDGVMEHPKCAAVKGTQVWENLFYNNVARKKTLQNSAD 192		
QY	182	DYKTYIDFSRAVHHIINTFSCRKIGANADYHASTSRDLAINSVGASVVDLIEI 241		
DB	193	DYKTYIDFSRAVHHIINTFSCRKIGANADYHASTSRDLAINSVGASVVDLIEI 252		
QY	242	KVSEDAADSIFFMDGYISNANVAKKTPMLIFINRLVQCTAKRAIEVYSAITPOAS 301		
DB	253	KVSEDAADSIFFMDGYISNANVAKKTPMLIFINRLVQCTAKRAIEVYSAITPOAS 312		
QY	302	KPEIYVSHLPSRVDVNIHPTKREVSLLNQERIETIRNAIEKILMSNTTPIFOQAL 361		
DB	313	KPEIYVSHLPSRVDVNIHPTKREVSLLNQERIETIRNAIEKILMSNTTPIFOQOK 372		
QY	362	NLSGIAQANPKQKVEASGSKTSOKIPVQSVMTDPNPGRGLHTYHHGSSNLEK 421		
DB	373	EYIQ-STLISQKSDSPVSKPQSKOKTQKVPVNMKVRDSSDPAGRLAFLQPKPSLPDK 431		

QY 422 FDLVS -VRNVRBRRNOKAGDLSRRELLVEIDSSFHGGLLDIYKNTCTYVGLADEAFAL 480
 DB 432 VSSLSVRSVRORRRPKETADLSVOELLAVDSCCHGMELEYRNTCTYGMADVAFAL 491
 QY 481 IOHNTRLVLYNVNVIKELMYOALCFGNFNAIOUSEPAPLOELLVMAKDEL -MSD 538
 DB 492 VOYNTHLYLVNVLNLSKELMYOQLTRFAFNAIOUSDPALELLALKEEDLDPGND 551
 QY 539 EKDEKLEIAVNTLEIKENAMINEFSTIHIDQCKLTRLVLYDQYTPMDRLPEFVL 598
 DB 552 TKDOLKRIAMNTLEIKEMLEEFVSHVHIDSSANLSRLPVLIDQYTPMDRVPFELL 611
 QY 599 ALGNDVYDWEDEKCFRIVASAVNGFYALHPILPNPSCNGIHLKKRNSMADEHENDL 658
 DB 612 CLGNDVWEDEKSCFOGVSAAIGNFYAMHPPLPNPSCDGIQFYSKRGSSQEKSDLEGN 671
 QY 659 ISDENVDQELLAEAAQAORENTIOHVLPFPMRLFLKPKKSMATDGTGVQVVASLEKLY 718
 DB 672 VDMEDNLDQDLSDAENMAQREWSIOHVLPFPMRLFLKPKKSMASNGTGVKVASLEKLY 731
 QY 719 KIFERC 724
 DB 732 KIFERC 737

RESULT 2

T51620
 DNA mismatch repair protein MLH1 (imported) - Arabidopsis thaliana (fragment)
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 03-Nov-2000
 C:Accession: T51620
 R:Jean, M.; Pelletier, J.; Hilpert, M.; Belzille, F.; Kunze, R.
 Mol. Gen. Genet. 262, 633-642, 1999
 A:Title: Isolation and characterization of AtMLH1, a MutL homologue from Arabidopsis the
 A:Accession: T51620
 A:Reference number: Z25418
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-737 <JEAN>
 A:Cross-references: EMBL:AJ012747; PIDN:CA010163.1
 C:Genetics:
 A:Gene: MLH1
 A:Map position: 4
 A:Superfamily: mismatch repair protein MLH1
 Query Match 67.5%; Score 2505; DB 2; Length 737;
 Best local similarity 66.4%; Pred. No. 5.5e-146; Indels 4; Gaps 3;
 Matches 482; Conservative 111; Mismatches 129;

QY 2 DEPSPPGGCAGRPRIIRLESVVRIRAGVIOFPSAVKELIENSIDAGASSVAV 61
 DB 13 EEPSPATTIVPREPKIORLESVVRIRAGVIOFPSAVKELIENSIDAGASSVAV 72
 QY 62 KDGLKLIQVSDGHRFEDLALICERTTSLAYEDLOTIKSGFGEALASMTYVG 121
 DB 73 KDGLKLIQVSDGHRFEDLALICERTTSLAYEDLOTIKSGFGEALASMTYVG 132
 QY 122 HVTYTTTTEGOLHGRVSYRVDGVMEHPKCAAVKGTQVAVENLFYNNVARKKTIQNSND 181
 DB 133 HVTYTTTTEGOLHGRVSYRVDGVMEHPKCAAVKGTQVAVENLFYNNVARKKTIQNSND 192
 QY 182 DYPKIVDFISRPVAVHINVTFSCKKHGANADVASSTSRDLAISKVYGA5VRLIEI 241
 DB 193 DYKIVDFISRPVAVHINVTFSCKKHGANADVASSTSRDLAISKVYGA5VRLIEI 252
 QY 242 KYSYEDADSIKMDGYISNNYAKKTIIMLFTINDRLVDTALKRIEIVYSAITLPOAS 301
 DB 253 KYSYEDADSIKMDGYISNNYAKKTIIMLFTINDRLVDTALKRIEIVYSAITLPOAS 312
 QY 302 KPFIYMSIHLPSHVDVNIHPKRVYSLNDRRIETRNAIEKILNMSVTRIFOTOL 361
 DB 313 KPFIYMSIHLPSHVDVNIHPKRVYSLNDRRIETRNAIEKILNMSVTRIFOTOL 372

QY 362 NLSGIAQANPQKDKVSEASMSGSTKQKIPYQWVRKIDPNDPNSGRILTYHAGSSNIEK 421
 DB 373 EYIQ -STLSQSDSVSGKPSGOKTQKVRVKNVKTIDSPAPRHLAFLOPKQSLPDK 431
 QY 422 FDLVS -VRNVRBRRNOKAGDLSRRELLVEIDSSFHGGLLDIYKNTCTYVGLADEAFAL 480
 DB 432 VSSLSVRSVRORRRPKETADLSVOELLAVDSCCHGMELEYRNTCTYGMADVAFAL 491
 QY 481 IOHNTRLVLYNVNVIKELMYOALCFGNFNAIOUSEPAPLOELLVMAKDEL -MSD 538
 DB 492 VOYNTHLYLVNVLNLSKELMYOQLTRFAFNAIOUSDPALELLALKEEDLDPGND 551
 QY 539 EKDEKLEIAVNTLEIKENAMINEFSTIHIDQCKLTRLVLYDQYTPMDRLPEFVL 598
 DB 552 TKDOLKRIAMNTLEIKEMLEEFVSHVHIDSSANLSRLPVLIDQYTPMDRVPFELL 611
 QY 599 ALGNDVYDWEDEKCFRIVASAVNGFYALHPILPNPSCNGIHLKKRNSMADEHENDL 658
 DB 612 CLGNDVWEDEKSCFOGVSAAIGNFYAMHPPLPNPSCDGIQFYSKRGSSQEKSDLEGN 671
 QY 659 ISDENVDQELLAEAAQAORENTIOHVLPFPMRLFLKPKKSMATDGTGVQVVASLEKLY 718
 DB 672 VDMEDNLDQDLSDAENMAQREWSIOHVLPFPMRLFLKPKKSMASNGTGVKVASLEKLY 731
 QY 719 KIFERC 724
 DB 732 KIFERC 737

RESULT 3

S43085
 DNA mismatch repair protein MLH1 - human
 N:Alternate names: mutL homolog 1 (hereditary nonpolyposis colorectal cancer)
 C:Species: Homo sapiens (man)
 C:Date: 20-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 29-Sep-1999
 C:Accession: S43085; 137414; 137549
 R:Bronner, C.E.; Baker, S.M.; Morrison, P.T.; Warren, G.; Smith, L.G.; Lescoe, M.K.; Kold, M.; Fishel, R.; Kolodner, R.; Liskay, R.M.
 Nature 368, 258-261, 1994
 A:Title: Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with
 A:Reference number: S43085; MUID:94195398; PMID:8145827
 A:Molecule type: mRNA
 A:Accession: S43085
 A:Residues: 1-756 <BRON>
 A:Cross-references: GB:007343; NID:9463988; PIDN:AA02085.1; PID:9463989
 R:Kolodner, R.D.; Hall, N.R.; Lipford, J.R.; Kane, M.F.; Morrison, P.; Finan, P.J.; B
 Cancer Res. 55, 242-248, 1995
 A:Title: Structure of the human MLH1 locus and analysis of a large hereditary nonpoly
 A:Reference number: 137414; MUID:95112274; PMID:7812952
 A:Accession: 137414
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-756 <RES>
 A:Cross-references: EMBL:U04978; NID:91079785; PIDN:AA82079.1; PID:91079787
 R:Papadopoulos, N.; Nicolaides, N.C.; Wei, Y.F.; Ruben, S.M.; Carter, K.C.; Rosen, C
 ation. P.; Lynch, H.T.; Peltoniemi, P.; Wecklin, J.P.; Chappelle, A.D.; Kinzler, K.W.;
 Science 263, 1625-1629, 1994
 A:Title: Mutation of a mutL homolog is associated with Hereditary Colon Cancer.
 A:Reference number: M49945; MUID:94174309; PMID:8128251
 A:Accession: 137549
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-218, 'V', 220-756 <RE2>
 A:Cross-references: EMBL:U07418; NID:9466461; PIDN:AAA17374.1; PID:9466462
 C:Genetics:
 A:Gene: MLH1; COCA2; HNPCC
 A:Cross-references: GDB:249617; OMIM:120436
 A:Map position: 3p21.3-3p21.3
 A:Introns: 39/2; 69/3; 102/3; 127/2; 151/3; 182/2; 196/3; 226/2; 264/1; 295/2; 346/3;
 C:Superfamily: mismatch repair protein MLH1
 C:Keywords: DNA repair
 Query Match 39.1%; Score 1452; DB 2; Length 756;

Best Local Similarity 39.6%; Pred. No. 2.7e-81;
Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;

```

OY 18 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 77
    8 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 67
OY 78 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 137
    68 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 127
OY 138 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 197
    128 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 187
OY 198 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 257
    188 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 244
OY 238 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 317
    245 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 304
OY 318 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 375
    305 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 363
OY 376 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 426
    364 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 421
OY 427 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 444
    422 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 481
OY 445 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 481
    442 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 481
OY 482 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 541
    542 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 601
OY 542 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 601
    602 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 661
OY 602 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 661
    662 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 693
OY 662 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 718
    694 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 750
OY 719 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 724
    751 IRRDETVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 756

```

RESULT 4

S54525 mismatch repair protein MLH1 - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein YMR520.16; protein YMR167w

C:Species: Saccharomyces cerevisiae

C:Date: 08-Jul-1995 #sequence revision 01-Sep-1995 #text change 29-Sep-1999

C:Accession: S54525; S48890; S54612

R:Hunt, S.; Bowman, S.

submitted to the EMBL Data Library, May 1995

A:Reference number: S54510

A:Accession: S54525

A:Molecule type: DNA

A:Residues: 1-769 <HUN>

A:Cross-references: GB:Z49705; EMBL:Z49700; NID:9825556; PIDN:CAA89803.1; PID:9825572
A:Experimental source: strain AB972
R:Protein, T.A.; Christie, D.M.; Liskay, R.M.
Mol. Cell. Biol. 14, 407-415, 1994
A:Title: Dual requirement in yeast DNA mismatch repair for MLH1 and PMS1, two homolog
A:Reference number: S48890; MUID:94088538; PMID:8264608
A:Accession: S48890
A:Molecule type: DNA
A:Residues: 1-257, 'L', 259-287, 'F', 289-707, 'L', 709-769 <PRO>
A:Cross-references: EMBL:U07187; NID:9460626; PIDN:AAAL6835.1; PID:9460627
C:Genetics:
A:Gene: SGD:MLH1
A:Cross-references: SGD:S0004777; MIPS:YMR167w
A:Map position: 13R
C:Function:
A:Description: DNA repair
C:Superfamily: mismatch repair protein MLH1
C:Keywords: DNA repair

Query Match 29.6%; Score 1097; DB 2; Length 769;

Best Local Similarity 32.8%; Pred. No. 1.7e-59;
Matches 265; Conservative 156; Mismatches 244; Indels 144; Gaps 19;

```

OY 17 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 76
    4 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 63
OY 77 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 136
    64 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 123
OY 137 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 196
    124 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 183
OY 197 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 255
    184 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 241
OY 256 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 314
    242 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 301
OY 315 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 374
    302 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 361
OY 375 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 411
    362 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 419
OY 412 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 442
    420 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 479
OY 443 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 471
    480 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 539
OY 472 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 539
    540 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 599
OY 530 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 579
    600 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 652
OY 580 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 639
    653 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 699
OY 640 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 695
    640 IRRLEESVYNNRIAGEVYQIORSAAVKELIENSLDAGASSVAVKDGKILQVSDGCH 695

```

Db 700 -----DMPKVDTSASISEPKAOFINRKHISLSLEHVLPCIKRR 742

QY 696 LKPKSMATDGTFOVASLEKTKIERC 724

Db 743 FLAPRHILKD--YVEIANLPDLKVFERC 769

RESULT 5

T50317
probable DNA mismatch repair protein, MHL1 homolog [imported] - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 02-Sep-2000
C:Accession: T50317
R:McDougal, R.C.; Rajandream, M.A.; Barrell, B.G.; Cadieu, E.; Lelaure, V.; Galibert, F. submitted to the EMBL Data Library, January 2000
A:Reference number: 225061
A:Accession: T50317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-684 <MCD>
A:Cross-references: EMBL:AL36536; PIDN:CAB66448.1; GSPDB:GN00067; SPDB:SPBC1703.04
A:Experimental source: strain 972h(-); cosmid cl703
C:Genetics:
A:Gene: SPDB:SPBC1703.04
A:Map position: 2
A:Insertions: 24/3; 70/3; 128/2
C:Superfamily: mismatch repair protein MHL1

Query Match 28.1%; Score 1040.5; DB 2; Length 684;
Best Local Similarity 32.7%; Pred. No. 4.3e-56;
Matches 241; Conservative 149; Mismatches 259; Indels 87; Gaps 13;

QY 17 RIRLESVVRNRIAGEVIOEPSAVKELIENSIDAGASSVAVKDGKLIQVSDGHC 76
Db 8 KIRPDOLVINKIAAGEIIRPENAIKELIENSIDAGSTSIDVLLKDGKLIQVSDGHC 67
QY 77 GIRPDLALICEHTTSKLSAYEDLOTIKSMGFRGALASMTYGVHVTYITTEGOLHGY 136
Db 68 GICVDDLPYLCORSTSKIDNFNDLOHLOTFGRGALASISVAKVTYITVLTSSDIHM 127
QY 137 RVSIRGCV-----MENEPKCAAVKQVAVENLFYNNVARKTTLONSNDYKTYDF 189
Db 128 KAFYVDALAPISPGMSPAPQPCGQGYTAEFLYNNRSKSLKNGSEFRIMTL 187
QY 190 ISRAVHINVTSCRHGANRADVHASSTSSRLDAIRSYGASVVRDLEIVSYEDA 249
Db 188 VOKAIHNDQVSEFCKKQVGDVSLSSISRLSKADKTRHLYGRVASHLDFSLGQSS 247
QY 250 DSIRKMGYISNANVYAKKTMILFINDRLVDCALAKRAIEFYASATLPQASKPIYMI 309
Db 248 IVGFANGFISNADQDKSNLILFINNRIVESVELHMLETYAKLHKGASVFLSL 307
QY 310 HLPSEHVNVNTHPTKREVSLLNOERITETIRNAIEEKLANSTTRIQOAL--NLGIA 367
Db 308 NNSPQDLVNVHPSKRIYHFLYDOELIATISCDKLGELIERDTRSPVLAQMLPSISNTR 367
QY 368 QANPKDKVSEKSGSKKIPVSOAVTDPNPS--GLHTYMGQSN-----LE 419
Db 368 NNE-----SSQAAVTYENTYLRTPRPRRSKSLSNFLQSSNNNDNEIE 416
QY 420 KKFEDLVANVRNRRN-----QKDAGLSSRHLLVELIDSSFRHGLLDIYKNCYV 471
Db 417 KYDSANNSNNATNDIKDLOTEIVEEGNSIDLESIKSLQKOVYNSMHVLAITNLEHKYV 476
QY 472 GLA--DEAFALIQHNRILVAVNISKEIWOALCFRGFNALITOLSEPAPOELIYMA 529
Db 477 GIVCPTRIAAVOAHNIGLVVDYGLSYHLFQICLTERGANGEVLEPLISIDLEIV 536
QY 530 LKDELSDEKDEKLEIAEVNTEILKENAMINETYSHIDODKTLRLPVLDQVYP 589
Db 537 NGDE-----DKSESEK-----TRILYSRDMIKDFYSISVSGGLTAVPMLSKTHP 586
QY 590 MDRLPEFVIAL--GNDVTMDKECFRTVASAVGNFYALHPILPSPSGNGLHLYKKNDS 648

Db 587 FEQPLLISSITKFPFMDKSCINGIMKAIKAF-----VPLP----- 626

QY 629 MADHARNLISDENDVDOLAEANAKQREMTIOHLYFSPMRFLKPKSMATDGT 708

Db 627 -----LSYESDVKSIRLS-----CLIEDYLFEPFRNRVLCPKKVFBEKCI 668

QY 709 VQVASLEKTKIERC 724

Db 669 YQITSLPLVNFERC 684

RESULT 6

T25389
hypothetical protein T28A8.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T25389
R:Lloyd, C. submitted to the EMBL Data Library, March 1997
A:Reference number: 220027
A:Accession: T25389
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-779 <MII>
A:Cross-references: EMBL:Z92813; PIDN:CAB07283.1; GSPDB:GN00021; CESP:T28A8.7
A:Experimental source: clone T28A8
C:Genetics:
A:Gene: CESP:T28A8.7
A:Map position: 3
A:Insertions: 16/1; 77/3; 149/1; 205/2; 241/3; 404/1; 515/3; 696/3
C:Superfamily: mismatch repair protein MHL1

Query Match 22.8%; Score 845; DB 2; Length 779;
Best Local Similarity 29.7%; Pred. No. 5.3e-44;
Matches 240; Conservative 136; Mismatches 276; Indels 156; Gaps 22;

QY 18 RIRLESVVRNRIAGEVIOEPSAVKELIENSIDAGASSVAVKDGKLIQVSDGHC 77
Db 25 IORLPQVNVNMAAGEVLNAPCAIKELIENSIDAGATEIMVMQNGKLIQVSDGHC 84
QY 78 IRPFDLALICEHTTSKLSAYEDLOTIKSMGFRGALASMTYGVHVTYITTEGOLHGY 137
Db 85 IERDPLVYCRFATSKLOKFEEDLMHMKTYGFRGALASLSHAKVAVYSKRADKAYO 144
QY 138 VSYRDGVNENPKCAAVKQVAVENLFYNNVARKTTLONSNDYKTYDFISRAVH 197
Db 145 ANFLDGKMTADTKPAAGKNGCTIATDFTNLPTRKMTTHGEBAKMNVDTLRPAHR 204
QY 198 INVPSCRHGANRADVHASSTSSRLDAIRSYGASVVRDLEIVSYEDAISFKMG 257
Db 205 PDVSAALQNOA--GDFPKDGDGNFRDVCNLLGRDVADELPLSL--NSTRIKFTTG 259
QY 258 VTSN-----ANYAKKTMILFINDRLVDCALAKRAIEFYASATLPQASKPIY 307
Db 260 HSKPIASATAIAONKRTSSEFSVFTNGRSVRODLKHPIDVLCAR--QLHAQCAL 317
QY 308 SIHSEHVNVNTHPTKREVSLLNOERITETIRNAIEEKLANSTTRIQOAL--NLGIA 367
Db 318 HLOIDETRIDVNVHPTKSVIFLEKELEIEIR--AYFEKVIQ-----EIFGEALDVKPE 372
QY 368 QANP-----QKDVSEA-----SMGSGTSOKIPVSOAVTDPNPS--GLHTYMGQSN 401
Db 373 EEOPTDIENLVMIPMSOLSKSIAIRKPDTPKPEKSPSAMSKKRDYMEVFRDAKER 432
QY 402 -----NPSGRL-----HTYMGQSSNLEKFFLVSVRNV-- 431
Db 433 IDEFVTRGCAVGTTSNDLIPGSGSLKRAKRTEDSGKEKEEDNTDFDVSMSVLSST 492
QY 432 ---RSRRNQKACD-----LSRHLELVIEDSSPHGLDVKNCYV 472
Db 493 ADGRLNESODLGEDDDVDVEYQKTHREFHPSIEVLRKREITANSOSIREKFKSTFVG 552

OY 473 LADEAFALIOHNTRLVNVNISKELATQALCREGNENAIOLSEAP-----LOELLYM 528
DB 553 SINWQVLLQFGTSLYHIDPSTVLAEEFYQISFSGNGLRDLDEPPILIEILL 610
OY 529 ALKDELMSEKDEKLEI-AEYNT-----EIKENAMENEXFISHIO--DGKL--FR 578
DB 611 ----GELSTRERPRVYAFEFVAVNENFPAEKLAEHADLHDHFALKDLOLENGRLHITE 666
OY 579 LPVLDQYTPDMRLPEFVALGNDVTWDEKCEPFTVASAVNGFYALHPPILPNSGNC 638
DB 667 ISLVAHYFVPOLEKLEPLFATLVAVNDYDEQNTFRTRICRAIGDLPTL----- 714
OY 639 IHHYKKNRSMADENHENDLISDEN--DVDQELAEAEANAOERWTIOHYLPFSRRL 636
DB 715 -----DTNFTTLDKKSIAFSATPW-----KTLIKVLMPLVKKRP 749
OY 697 KPPKSMATDGTQFVQVSLKLYKIFERC 724
DB 750 IPPEHFKQAGVIRQLADSHDLKVFERC 777

RESULT 7

DNA mismatch repair protein [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence-revision 10-May-2001 #text-change 22-Oct-2001

C:Accession: C89904

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: C89904

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-669 <KUR>

A:Cross-references: GB:BA000018; PID:g13701096; PIDN:BA042391.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: mult

C:Superfamily: mismatch repair protein hex

Query Match

Best Local Similarity 16.1%; Score 597; DB 2; Length 669;

Matches 174; Conservative 123; Mismatches 216; Indels 128; Gaps 18;

OY 17 RIRLEESVYVNRVIAAGEVIOQRPSSAVKELIENSLDAGASSVAVKDGKLIQVSDGH 76
DB 3 KIKELQTSIAKKAAGEVERPSSVYKELLENAIDAGATEISIEVESGVQSIKYVNDGS 62
OY 77 GIRFEDLALICERHTSKLSAYEDLOTIKSMGRGALASMTYGVHTVTTITTEGOLHGY 136
DB 63 GIEAEDLGLVHRHATSKLDQEDDLPHIRLGRGALASISSAVATTLKTCID-NANGN 121
OY 137 RVSRYDGVMENEKPCAAVAGTQVAVENLFYNNVARKKTLONSNDYPRKIVDISRFVAV 196
DB 122 ELYVENGELINH-KPAKAKKAGTDLVESLFYNTPARLKYIKSLYTELGKTTDIVNNMAM 180
OY 197 HINVTESCKRHGARNADVHASTSSRLDAIRSYGASVVRDLKIVSYEDADSIKMD 256
DB 181 HPIRITALISDGTMLSTSGRTN--EVAEYIGKMKVADLVHIS--GDTSD-YHIE 233
OY 257 GYI-----SNAYVAKRTIMLFINDRLVDTALRAIEFYVATLPQASKPTIYMSI 309
DB 234 GFVAKREHSRKNHYIS-----IFNGRYIKFMKMLKAIIEGYHLLTIGRPICYN 287
OY 310 HLPSEVAVNIHPTKVEVSLNORLIEITRNAIEEKLKNSNTRIFQOAL-----NLG 365
DB 288 EMPPIIVAVNVHPTKLEVLSEKQOLYQLIVSKIQ-----AFKDRILLPKNLDY 338
OY 366 IAOANPOKQVSAKSGSGTKQKIPVSOQVNRDPNPSGRHLTHYHGSSNLEKKFDLY 425

DB 339 V-----EKKKRVLYS-----PEQKIEREQONTENNOEKTFSSSESNKSFMAENONDEI 369
OY 426 SVRN-----VRSRRN-----QKDACGLSSRHLELVIEDSPHGLDIY 465
DB 390 VIEDSDYNPVTSTSLITDESSGYNNRREDEDEFFKQOELQEMDQTFPSNEDASV 449
OY 466 KN-----CYVGLADEAFALIOHNTRLVNVNISK 496
DB 450 QNTEKASDDYVDVNDIKGTSKDPKRIPIYMEIVQGVHGTYIIAQNEFQMYMIDQAAQ 509
OY 497 KELMYQALCRFQNFNFIQSEAPQOELLYMAL-----KDELMSEKDEKLEIA----- 548
DB 510 ERIKYEYFRDKIGEV-----NEVQDILLPTLFHESKDEQLVTDQKNELOQVIMLE 562
OY 549 -----EVNTEIKENAMENEXFISHI 570
DB 563 HFGCHDIYVSYVWFPKDEVE-EIKMDLIEILLKVKYDI 602

RESULT 8

DNA mismatch repair enzyme, Mult. [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence-revision 14-Sep-2001 #text-change 30-Sep-2001

C:Accession: E97126

R:Nolling, J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: E97126

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-622 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK79800.1; PID:g15024811; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC1836

C:Superfamily: mismatch repair protein hex

Query Match

Best Local Similarity 15.8%; Score 584.5; DB 2; Length 622;

Matches 170; Conservative 128; Mismatches 262; Indels 67; Gaps 17;

OY 17 RIRLEESVYVNRVIAAGEVIOQRPSSAVKELIENSLDAGASSVAVKDGKLIQVSDGH 76
DB 2 RIMLESDTSNKIAAGEVERPSSVYKELLENSIDKAKTINIETENGRTLLKVLDDGY 61
OY 77 GIRFEDLALICERHTSKLSAYEDLOTIKSMGRGALASMTYGVHTVTTITTEGOLHGY 136
DB 62 GIDKDIKAFMHPHATSKISKLDIYSINTLGRGALPSIAGSVKTLKSRKNEFG 121
OY 137 RVSRYDGVMENEKPCAAVAGTQVAVENLFYNNVARKKTLONSNDYPRKIVDISRFVAV 196
DB 122 EISISGSVD-YIKDCGTNIGHIEVADLFYNNVARKKFLKSTAKEASISDLYNRLALA 180
OY 197 HINVTESCKRHGARNADVHASTSSRLDAIRSYGASVVRDLKIVSYEDADSIKMD 256
DB 181 HSEISFRLNNG--KVITTYATDNLIDTRALTYGKKICCNVI-----SFRHND-LVSVH 233
OY 257 GYISNANYV-AKRTIMLFINDRLVDTALRAIEFYVATLPQASKPTIYMSIHPSEH 315
DB 234 GYVNAEISRSGSRNNOISIFINKRYIKKLTITAVENAVKSFMMINKPFFITFDIPPE 293
OY 316 VDVNHPHTKVEVSLNORLIEITRNAIEEKLKNSNTRIFQOAL-----NLGIAQAN 370
DB 294 VDVNHPHTKVEVFONERDLFKIIFDVHIGIRNS-LKESFEALKEEKEDKLFEDKEDV 352
OY 371 POKDKVSEASMGSG-TRSKQIPVSOQVNRDPNPSGRHLTHYHGSSNLEKKFDLYSVN 429
DB 353 IYKNEIKHDKIDGQIYIKKDSFPVQPIPLDKRP-----IENYDSTKENKRSDDREKD 409
OY 430 VVRSRRNQKADGLSSRH-----ELLVEIDSFHPGLDIVKNCYVGLADEAFALIOH 484

Db 410 IIKETSOEKETEDYDITTKKAKPELRY-----IGOFNNYTLIAESF 450
OY 485 TRLYVNVVNVNISKELMYOALCRFGNFAIOLSEPAFLDELLMALKDDELSDEKDK 544
Db 451 EELVITIDQHAHEKLEFEKREDIKKNG---VSQILLTPSVVELLPEDFIYDE-----502
OY 545 LEIAEVNTEILKENAEMINEYF---SIHIDODGKTRLPVLLDOYPPMDRLPEFYLAIG 601
Db 503 -----NKEVER-NAGEVIEFGDNTVAIKVPLFLGKPLVKDLLEIIDLNKN---MG 551
OY 602 NDVTMDDEKECFRIVA--SAVGNFYAL 626
Db 552 SGETSVKYSRATACSAVKAAYHEL 578
RESULT 9
A:Accession: A95020
DNA mismatch repair protein hexb [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: A95020
R:Jettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umeyam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
son, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: A95020
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-649 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74354.1; PID:g14971640; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0173
C:Superfamily: mismatch repair protein hexb
Query Match 15.2%; Score 565; DB 2; Length 649;
Best Local Similarity 27.7%; Pred. No. 6.2e-27;
Matches 162; Conservative 115; Mismatches 219; Indels 88; Gaps 16;
OY 18 IRLRESVNRVRIAGEVIOEPSAVKELIENSIDAGASSVAVAVKDGGLKLIQVSDG 77
Db 4 IIEPEMLANOIAGEVIERPAVKEVLENALDAGSSQIIIEIEAGIKKVOITDNGH 63
OY 78 IREEDLALICERHTTSKLSAYEDLOTIKSMGFRGALASMTYVGHVYTTTTEGOLHGR 137
Db 64 IAHDEVELALRRHAKTSIKKNOADLEFRITLGRGELPSIASVSYLITLITAVDASHG 123
OY 138 VSYRDGVNENPKCAVAGTQVAVENILFYNNVARKKTIONSDYPRKIVDFISRAVHH 197
Db 124 LVARGGEVE-EVLPATSPGTVKCEVDFENTPARLKTKYKSOALSHIIDIVNRLGLAH 182
OY 198 INVTFSCRRKGANRADVHASSTSRDLAIRSVYASVVRDLIEIKVSTEDAADSTFKMDG 257
Db 124 LVARGGEVE-EVLPATSPGTVKCEVDFENTPARLKTKYKSOALSHIIDIVNRLGLAH 182
OY 198 INVTFSCRRKGANRADVHASSTSRDLAIRSVYASVVRDLIEIKVSTEDAADSTFKMDG 257
Db 183 PELSFSLISDG--KEMTRTAGTQGLROHAGITGLVSAKKMIEIENS-----DLDFEISG 235
OY 258 YIS-----NANYVAKKITMLFINDRLVDCALRAIEFYVSATLPQASKPIFYMSIH 310
Db 236 FVSLPELTPRANRNYIS-----LFINGRYIKNPLNRLNAILDGFSGKLMVGRPLAVIHH 289
OY 311 LPSEHDVNIHPTKEVSLINOERII-----ETIRNAIEE-----KLANSNTR 354
Db 290 IDPYLADVNVHPTKQVARIKKEKELMTLVSEALANSLEKOTLLPDALLENLAKSTVNRREK 349
OY 355 IFQT-----QALNLSGIAOANPQKDYSEASMGSGTKSOKIYVSOAVRTPDPNPSGRHH 408
Db 350 VEOITLPLKENTLYEKEPESRPSQTEVADYVELTDEGDLTFAKETIDRLTKRAKHH 409
OY 409 TYWHGQSSNLEK-----KFDIVSVRVNVRSRNOKDAGDLSRHELLVEIDSSFPGLDIT 464

Db 410 -FAERKPANYDQIDHPELDIASI-----DKAYDKLEREASPEL---449
OY 465 VKNCTYGLADEAFALIOHNTRLVNVVNVNISKELMYOALCRFGNFAIOLSEPAFLDE 524
Db 450 -----EFFGQNHGTYLFAQGRDGYIIDQHAQBERVYEEYRESISNVDQSO-----QQ 498
OY 525 LVNAL-----KDELSMDEK-----DDEKLEIAEV--NTEILEKN 558
Db 499 LVPIYIEFPADALRLKEMPLIEVGYLAEYGENQFLREH 542

RESULT 10
A:Accession: H97891
DNA mismatch repair protein [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: H97891
R:Hoekings, J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; Mohren, S.
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: H97891
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-649 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK98964.1; PID:g15457702; GSPDB:GN00174
C:Genetics:
A:Gene: hexb
C:Superfamily: mismatch repair protein hexb

Query Match 15.2%; Score 565; DB 2; Length 649;
Best Local Similarity 27.7%; Pred. No. 6.2e-27;
Matches 162; Conservative 115; Mismatches 219; Indels 88; Gaps 16;
OY 18 IRLRESVNRVRIAGEVIOEPSAVKELIENSIDAGASSVAVAVKDGGLKLIQVSDG 77
Db 4 IIEPEMLANOIAGEVIERPAVKEVLENALDAGSSQIIIEIEAGIKKVOITDNGH 63
OY 78 IREEDLALICERHTTSKLSAYEDLOTIKSMGFRGALASMTYVGHVYTTTTEGOLHGR 137
Db 64 IAHDEVELALRRHAKTSIKKNOADLEFRITLGRGELPSIASVSYLITLITAVDASHG 123
OY 138 VSYRDGVNENPKCAVAGTQVAVENILFYNNVARKKTIONSDYPRKIVDFISRAVHH 197
Db 124 LVARGGEVE-EVLPATSPGTVKCEVDFENTPARLKTKYKSOALSHIIDIVNRLGLAH 182
OY 198 INVTFSCRRKGANRADVHASSTSRDLAIRSVYASVVRDLIEIKVSTEDAADSTFKMDG 257
Db 183 PELSFSLISDG--KEMTRTAGTQGLROHAGITGLVSAKKMIEIENS-----DLDFEISG 235
OY 258 YIS-----NANYVAKKITMLFINDRLVDCALRAIEFYVSATLPQASKPIFYMSIH 310
Db 236 FVSLPELTPRANRNYIS-----LFINGRYIKNPLNRLNAILDGFSGKLMVGRPLAVIHH 289
OY 311 LPSEHDVNIHPTKEVSLINOERII-----ETIRNAIEE-----KLANSNTR 354
Db 290 IDPYLADVNVHPTKQVARIKKEKELMTLVSEALANSLEKOTLLPDALLENLAKSTVNRREK 349
OY 355 IFQT-----QALNLSGIAOANPQKDYSEASMGSGTKSOKIYVSOAVRTPDPNPSGRHH 408
Db 350 VEOITLPLKENTLYEKEPESRPSQTEVADYVELTDEGDLTFAKETIDRLTKRAKHH 409
OY 409 TYWHGQSSNLEK-----KFDIVSVRVNVRSRNOKDAGDLSRHELLVEIDSSFPGLDIT 464

QY 525 LVMAL-----KDELMASDEK-----DDEKLEIAEV--NTEILKEN 558
DB 499 LVPYIFERPADDALRLKERMPLEEVGFVLAEGYGENOFILREH 542

RESULT 11

A33589 mismatch repair protein hexb [validated] - Streptococcus pneumoniae
C:Species: Streptococcus pneumoniae
C:Date: 27-Feb-1990 #sequence_revision 27-Feb-1990 #text_change 26-May-2000
C:Accession: A33589
R:Rudhomme, M.; Martin, B.; Mejean, V.; Claverys, J.P.
J. Bacteriol. 171, 5332-5338, 1989
A:Title: Nucleotide sequence of the Streptococcus pneumoniae hexb mismatch repair gene:
A:Reference number: A33589; MUID:90008767; PMID:2676973
A:Accession: A33589
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-649 <PRD>
A:Cross-references: GB:M29686; NID:9133656; PIDN:AAA88600.1; PID:9133657
C:Function:
A:Description: required for mismatch repair [validated, MUID:90008767]
C:Superfamily: mismatch repair protein hexb

Query Match 15.1%; Score 560; DB 2: Length 649;
Best Local Similarity 27.6%; Pred. No. 1.3e-26;
Matches 161; Conservative 115; Mismatches 220; Indels 88; Gaps 16;

QY 18 IRRLSESVNRIAGAEVIOEPSAVKELIENSIDAGASSVAVKGGGLKLIQVSDGH 77
DB 4 IIEPLMALNOIAGAEVIERPASYCKEVENALDAGSSQIIIEELGKKVQITDNGH 63
QY 78 IRREDLAIICERHTTSKLSAYEDLOTIKSGFGEALASMTYVGHVYTTTTEGOLHGY 137
DB 64 IAHDEVALLRRHATSKINKNOADLFRIITLGEALPISASVYLTLTAVGASHGTR 123
QY 138 VSTRGVMENEPKCAAVGTOVMENLFYNNVARKKTLONSNDYPKYVDSIRAVH 197
DB 124 LVAKEGEVE-EVLPATSPVTKCVEDLFENFPAKLYKSOQAEISHIIDYINRLGLAH 182
QY 198 INTVFSCRHGANRADVHSASTSSRLDAIRSYGASVNRLEIKVSYEDADSIKMG 257
DB 183 PEISFSLISDG--KEMTRAGTGOLRQALAGIYGLVSAKMIIEENS-----DLDEISG 235
QY 258 YIS-----NANYVAKITMTLIFINDRLVDTALRAIEFYASATLPQASRPFYMSI 310
DB 236 FVSLPELFRANRYIS-----LFINGRYIKNPLNRAILIDGSKLMGRPFLAVIH 289
QY 311 LPSEHVDVNIHPKKEVSLNDEKIT--ETRNAIEE-----KLMSNTTR 354
DB 290 IDPYLADVNVHPKQEVRISEKELMTLVSEALANSIKEDTLIPDLLENLAKSTVRNREK 349
QY 355 IFOT-----QALNLSGIAOANPOKDVSEASGSGTKSOKIPVSGMVRTDPNPSGRH 408
DB 350 VEDTILPKEKNTLYEKTEPSRSQTEVADYQVLEDEGODTLFAKETIDRLTKRAKHA 409
QY 409 TYWHGGSNLEK-----KFLDVSVRNVRSRNOKADGLSSRHLLVEIDSSPHGLDI 464
DB 410 -FAERKPVANDQDLHDELDAI-----DKAVDKLEREASFPPEL--- 449
QY 465 VKKCTVGLADEAFALIOHNTRLYLIVNVNISKELMYOQALCFGNFATQISEPAPLOE 524
DB 450 -----EFFGGMHGTLYLFAOGDGLYIIDHAAQERVYEEFERESIGVNDGSO-----QQ 498
QY 525 LVMAL-----KDELMASDEK-----DDEKLEIAEV--NTEILKEN 558
DB 499 LVPYIFERPADDALRLKERMPLEEVGFVLAEGYGENOFILREH 542

RESULT 12

H86900 DNA mismatch repair protein mutL [imported] - Lactococcus lactis subsp. lactis (strain
C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: H86900
C:Species: Lactococcus lactis
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: E64046
R:Flaeschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Keriavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Status: preliminary
A:Accession: H86900
A:Molecule type: DNA
A:Residues: 1-695 <STO>
A:Cross-references: GB:AE005176; PID:g12725273; PIDN:AAK06306.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: hexb
C:Superfamily: mismatch repair protein hexb

Query Match 15.0%; Score 555; DB 2: Length 695;
Best Local Similarity 30.7%; Pred. No. 2.8e-26;
Matches 178; Conservative 99; Mismatches 237; Indels 66; Gaps 17;

QY 17 RIRLESVNRRIAGAEVIOEPSAVKELIENSIDAGASSVAVKGGGLKLIQVSDGH 76
DB 42 KIIEMLBALANOIAGAEVIERPASYCKEVENALDAGSSQIIINVERAGRLIEVDNGL 101
QY 77 GIRPEDLAIICERHTTSKLSAYEDLOTIKSGFGEALASMTYVGHVYTTTTEGOLHGY 136
DB 102 GLEKEDEVALLRRHATSKINKNOADLFRIITLGEALPISASVQMTITSTAEESGT 161
QY 137 RVSTRGVMENEPKCAAVGTOVMENLFYNNVARKKTLONSNDYPKYVDSIRAVH 196
DB 162 KLVAKGNIET-LEPLAKRVGTRISVANLEFNPAPKLYKSOQAEISHIIDYINRLSLA 220
QY 197 HINVFSCRHGANRADVHSASTSSRLDAIRSYGASVNDLIEIKVSYEDADSIKMG 256
DB 221 HPEISFLVNG--KEPLKTAGNGDLRQVLAITYGISTAKMKQIK-----GSDLPELT 273
QY 257 GYIS-----NANYVAKITMTLIFINDRLVDTALRAIEFYASATLPQASRPFYMSI 309
DB 274 GYSLPELFRANRYIS-----ILINGRIFKNPLNRAILIDGSKLMGRPFLAVI 327
QY 310 LPSEHVDVNIHPKKEVSLNDEKIT--ETRNAIEEKLMSNTTRIFQYAL- NLSGIAO 368
DB 328 KIDPTLADVNVHPKQEVRISEKELMTLVSEALANSIKEDTLIPDLLENLAKSTVRNREK 382
QY 369 ANPOKDVSEASGSGTKSOKIPV-SQMVTRDPNPSGRHLYTYWHGGSN--LEKKE--- 422
DB 383 -----KEKES-LVQTEPLRQNNPLLYDNRQDYVREAFKINKNOASDSEQTDFN 437
QY 423 -----DLVSVNRNVRSRNOKADGLSSRHLL-VEID-----SSFHGLDIY 465
DB 438 TENQMTDAVSEKMTGRTVSTTEITDINSLENTVSNFEIDFNDKATSGSTFEPQL--- 493
QY 466 KNTYVGLADEAFALIOHNTRLYLIVNVNISKELMYOQALCFGNFATQISEPAPLOE 525
DB 494 ---EYLAQHLATYLLCCSGEGLYLVDOHAAQERIKTEYWKDKIGEVSMEOIILAPY--L 548
QY 526 LVMALKDELMASDEKDEKLEIAEVNTEILKENAMEMINEY 565
DB 549 FTLAKNDPFLAKKD--LLHAGVFLIEEGYGENOFILREH 566

RESULT 13

E64046 mismatch repair protein mutL - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: E64046
R:Flaeschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Keriavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MID:95350630; PMID:7542800
 A:Accession: E64046
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-629 <TIGR>
 A:Cross-references: GB:032692; GB:142023; MID:91573013; PIDN:AAC21745.1; PID:91573016;
 C:Genetics:
 A:Gene: mult
 C:Superfamily: mismatch repair protein hexb
 C:Keywords: DNA repair

Query Match 14.8%; Score 547.5; DB 2; Length 629;
 Best Local Similarity 27.3%; Pred. No. 7e-26;
 Matches 174; Conservative 119; Mismatches 251; Indels 93; Gaps 19;

18 IRRLEESVYVRIAGVYIORSAYKELIENSIDAGASSVAVAKDGLKIQVSDG 77
 3 IKILSPOLNQAAGVYVERPASYKELIENSIDAGANKIDIDENGANLIRIDMGCG 62
 78 IRRLEESVYVRIAGVYIORSAYKELIENSIDAGASSVAVAKDGLKIQVSDG 137
 63 IRRLEESVYVRIAGVYIORSAYKELIENSIDAGASSVAVAKDGLKIQVSDG 197
 138 VSYRQGVMEPEPCAAVGTQVMEVNFYVNVARKKTLQNSNDYPRKIVDFISFVAVH 182
 123 VYAGGRMETTIRPASPVGITVEANLFPNPARKEFLRTKTESHIDEVIRIATLK 256
 198 INVPSCRRKANRADVHSA-STSSRLDAIRSYGASVVRDLIELKVSIEDAADSIFKMD 237
 183 ENFAPFLTHNKILROYRPAEALIMQOLKRVAAICGDFYKALRIKWKHD-----LHLS 237
 257 GYISNANYAKITM-IIFINDRLVDTALKRAIEFYVSATLPQASKPIYMSIHLPSEH 315
 238 GWVAPNPSRQNDLSTCYINGRVRKQVISHAIROAYQYLPTDAVPVAFLEIDNPMD 297
 316 VDVNHTPTKEVSLNOERIIETIRNAIEKLNSTRTIRFOQALNLSGIAQNPQDK 375
 298 VDVNHTPTKEVSLNOERIIETIRNAIEKLNSTRTIRFOQALNLSGIAQNPQDK 351
 376 VSE-----ASMGSGT-----KSOQIPVSOVMTDRPNPGRGLHYHGGSS 416
 352 VREPOPNYSIRPNRATGNSAPQYHKKPQONOPHFNTPMFNHSTGDRSDAPS 411
 417 NLEKKIDLVSVNVRSRRNQDAGDLSRH-----ELVEIDSPFHPILDIVK 466
 412 KTEGRVYAEIIRLPPTA--QKDISPTAQONISDTAKIISREI-----IECSHLLALSLIE 466
 467 NCTIYGLADAFALIOHNTRELYLVNVNISKEIMTQOALCRFGNFIQLEPAPIDEL 526
 467 N-----RALILOONODFLLSLEKLOL-----KOQIDIEQALLLP-I 509
 527 VMALKDELMSEK--DDEKLEIAEVNTEILKENAMINEYFSIHIDDGKLT--RLPV 582
 510 VFRLESQFOQMOQYSDFK-----KIGFEFIEHQADL-----RLTKRVPSA 552
 583 LDQYTPMDRLPEFVALAGNDVTWDEKECFRTVASA 619
 553 L-----RTQNLQKVMAM--LFRDENSSPITALCA 581

RESULT 14

C181860
 DNA mismatch repair protein NMA1655 [imported] - Neisseria meningitidis (strain 22491 se
 C:Species: Neisseria meningitidis
 C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C:Accession: C81860
 R:Parthill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churche, C.; Klee, S.R.; Morel
 ; Holroyd, S.; Jagers, K.; Leather, S.; Mout, S.; Mungall, K.; Quail, M.A.; Rajandream,
 Nature 404, 502-506, 2000
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.
 A:Reference number: AB1775; MID:20222556; PMID:10761919
 A:Accession: C81860
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-658 <PAR>
 A:Cross-references: GB:AL162756; GB:AL157959; MID:97380091; PIDN:CA848483.1; PID:9738
 A:Experimental source: serogroup A, strain 22491
 C:Genetics:
 A:Gene: mult
 C:Superfamily: mismatch repair protein hexb

Query Match 14.5%; Score 538; DB 2; Length 658;
 Best Local Similarity 25.7%; Pred. No. 2.9e-25;
 Matches 172; Conservative 122; Mismatches 246; Indels 128; Gaps 20;

18 IRRLEESVYVRIAGVYIORSAYKELIENSIDAGASSVAVAKDGLKIQVSDG 75
 2 PRIALPHLNOIAGVYVERPAMALKEIENSIDAGATAIDVEIDGGIHLIRVSDMG 61
 76 HIREDEALICERTTSKLSAYEDLOTISMGFRGALASMYGVHVTITTEGOLHG 135
 62 SIHPDDIEALHRAVSKITLNDLEHVASMGFRGALASMYGVHVTITTEGOLHG 121
 186 VSYRQGVMEPEPCAAVGTQVMEVNFYVNVARKKTLQNSNDYPRKIVDFISFVAVH 195
 122 TOVKADGKL--SSPTAAHPVGTTEAELEFNTPARKEKSENEYAHCAVTERLAL 180
 196 HINVTPTSCRRKANRADVHSA-STSSRLDAIRSYGASVVRDLIELKVSIEDAADSIFKMD 255
 181 AHPIHAFSLKRGKQVFKLPASLHERI-----AAVGDQDQTSLEIDSGCALRL 232
 256 DQYISNANY-ARKITMILFINDRLVDTALKRAIEFYVSATLPQASKPIYMSIHLPSE 314
 233 YGAIAPPTPKAGTKDQCYFVNHREFVRDKVMLHAKOAYRVYLHNAITPAFVLDLPE 292
 315 HDVNHPTKEVSLNOERIIETIRNAIEKLNSTRTIRFOQAL-----NLGI 366
 293 AVDVNHTPTKEVSLNOERIIETIRNAIEKLNSTRTIRFOQAL-----NLGI 349
 367 AQA-----NPOKDVSEAS-----MGSTKQKIPVSOVMTDRPNPGRGLHYHGGSS 409
 350 TPAPNPSRQNDLSTCYINGRVRKQVISHAIROAYQYLPTDAVPVAFLEIDNPMD 297
 407 ---LHTYHGGSSNLEKKIDLVSVNVRSRRNQDAGDLSRH-----ELVEIDSPFHPILDIVK 466
 410 RAAMNTY-----ALYKTDIDLE--LSQFOARFGNMPSETP-APTDPPLSDGIPS 460
 461 ---LHTYHGGSSNLEKKIDLVSVNVRSRRNQDAGDLSRH-----ELVEIDSPFHPILDIVK 500
 461 OSELPPLGPAIALLGIY-----ILAQEDSLILDMAAAEVNYEKMKROENGRIOS 516
 501 -----YQOALCRFGNFIQLEPAPIDEL-----PAPIDELIYV 528
 517 ORLLIPTFAASHERCALADYAEIAGFG-----LELSDMGNTLAVRAVPTMIGADV 572
 529 ALKDDLELMSEK--DDEKLEIAEVNTEILKENAMINEYFSIHIDDGKLT--RLPV 586
 573 SLARVITG-----ELAQVSSQTEIEHNLATMTSCGVSVRAGROLTLP-ENNAL 622
 587 TPDMDRLP 594
 623 LRDMENTP 630

RESULT 15

AH1615
 DNA mismatch repair protein [imported] - Listeria innocua (strain C11p1262)
 C:Species: Listeria innocua
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
 C:Accession: AH1612
 R:Glasger, P.; Fingand, L.; Buchteser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussungel, O.; Entian, K.D.; Esht
 ; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapatk, G.; Madueno, E.; Maitournam, A.
 A:Reference number: AB1775; MID:20222556; PMID:10761919
 A:Accession: C81860
 A:Status: preliminary

A:Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AH1612
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-603 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC96672.1; PID:g16413914; GSPDB:GND0178
A:Experimental source: strain C11P11262
C:Genetics:
A:Gene: mult.
C:Superfamily: mismatch repair protein hexB

Query Match 14.3%; Score 532; DB 2; Length 603;

Best Local: Similarity 26.6%; Pred. No. 5.9e-25; Matches 173; Conservative 109; Mismatches 237; Indels 132; Gaps 19;

```
OY 18 IRLLESVVRIRIAGEVYQIPSSAVKELIENSIDAGASSVAVKDGLKLIQVSDGHC 77
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 5 IVELTDLAKSKIAAGEVERPASYKELVEMALDAGSTVIDIIVEAGLKKITIIDNGSG 64

OY 78 IREDDALICERTTSKLSAYEDLOTIKSGFGEALASMTYVGHVVTITTEGOLHGYR 137
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 65 IEEEDVAIAFLRHATSKIKHEADLFRVHTLGFGEALPSIASVSHLTLEST-GETKGT 123

OY 138 VSTRDGVMEPRPCAAVKTQYVENVLFYNNARKKTLONSDDYPKIVDFISRAVHH 197
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 124 ISLEGKI-TEOKSGHARKGTQIEVSOLFNTFARLRYLKLPLPELGNITDILNRLALAH 182

OY 198 INTFOSCKHGANRADVHASTSSRLDAIRSYGASVVRDLIEIKVSEDAADI-FKMD 256
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 183 PDISFRSHNG--KPLQITNGNDLROVIAITIGVSIARSKIPIK-----AESLDKIS 234

OY 257 GYI-----SNANYAKKITMLFINDRLVDTALRAIEFYSAVLPOASKPIFYMSI 309
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 235 GYAVLPEVNSNRNYS-----TIINGRFIKFPALVKAIOEGYHTLLPIGRPIIYLOI 288

OY 310 HLPSEHDVNIHPTKKEVSLINQERILETRNAIEEKLMSNTTRIFOTQALNLSGIAQA 369
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 289 EMPDIIVDVAVHPAKLEVRLSKEKELGOLISOMIKD-----AFHQQLIPDGEVSKR 340

OY 370 NPQDKVSEASMGSGTFSOK-----IPVSQMV-----RT 398
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 341 QKEQOKSEQIQMSFEERKQPKPEPTLFSKNPIIEYVPSDEIVREDDFILEMPTYNPET 400

OY 399 DPRN-----PSGRIL-TYMHGSSNLEKKFDLVSRRNVVRSRRNOKDAGDLS 445
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 401 OPEQVEOKERIPKMPYPIGOMHATYIFPAQENGLYIIDQAQERIKYEFYREKIGEVSR 460

OY 446 R-HELLVEIDSSFPGLDIYKNCITYGLDEAFALIQHNTRLYLNVVNIISKELMTQA 504
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 461 ELQELLVPIVLEPP-----SDEVYRLIEOKAKLEEVGVF-----494

OY 505 LCRRGNFNATQLESPAPLOELVWALKDDEMDKDEKLEIAEVTEILKENAEMI-- 562
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 495 LENFGNSFTIRHPTWFK-----DQEEMLREIIDEALSAAPSIHKLREPTALMS 548

OY 563 -----NEFYSIHIDQDGKTLRL-----PVVLDQYTPDMRL 593
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 549 CKKSIRKANHLYTMQ-DWEALDLTLREASDPFTCHGRPVIIQYSTYELKKM 598
```

Search completed: March 27, 2003, 15:45:36
Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.D5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 15:42:15 ; Search time 42 Seconds

(Without alignments)
3551.860 Million cell updates/sec

Title: US-09-954-950-2

Perfect score: 3709
Sequence: 1 MDESPRGGCGACGPEPRIR.....DTFVQVASELEKTYIFERC 724

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organelle:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*
15: SP-virus:*
16: SP-bacteriophage:*
17: SP-archaeal:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3645	98.3	715	10	O8S2D3	O8S2d3 oryza sativ
2	2505	67.5	727	10	O9C5F9	O9C5f9 arabidopsis
3	2505	67.5	737	10	O9ZRY4	O9Zry4 arabidopsis
4	1438	38.8	760	11	O8VDY4	O8vdy4 mus musculus
5	1261	34.0	664	5	O9V380	O9v380 drosophila
6	1255.5	33.9	663	5	O61917	O61917 drosophila
7	1040.5	28.1	684	3	O9P7W6	O9p7w6 schizosach
8	911.5	24.6	887	5	O9BIX4	O9bix4 trypanosoma
9	845	22.8	779	5	O9XU10	O9xu10 caenorhabd
10	788	21.2	484	5	O8TON1	O8ton1 drosophila
11	684.5	18.5	484	11	O9CTA7	O9cta7 mus musculu
12	580.5	15.7	468	2	O9LAP6	O9lap6 staphylococ
13	553	14.9	563	5	O8SS00	O8ss00 cephalothic
14	539.5	14.5	643	16	O8RG56	O8rg56 fusobacteri
15	530	14.3	590	16	O8RA70	O8ra70 thermococ
16	524.5	14.1	656	17	O8TBS5	O8tbs5 methanosarc

17	513.5	13.8	323	2	O93R03	O93R03 lactococcus
18	512.5	13.4	374	4	O9BQEO	O9BQEO homo sapien
19	461	12.4	628	16	O8UH13	O8UH13 agrobacteri
20	429.5	11.6	923	10	O94116	O94116 arabidopsis
21	420	11.3	899	5	O9V7B6	O9V7B6 drosophila
22	414.5	11.2	908	3	O8RG50	O8RG50. saccharomyc
23	414.5	11.2	908	3	O8RG48	O8RG48. saccharomyc
24	414	11.2	895	5	O8T9C0	O8T9C0 drosophila
25	411	11.1	526	2	O9A155	O9A155 burkholderi
26	405	10.9	805	5	O9TVL8	O9TVL8 caenorhabdi
27	402	10.8	779	10	O81287	O81287 arabidopsis
28	401.5	10.8	908	3	O8RG57	O8RG57 saccharomyc
29	400.5	10.8	908	3	O8RG54	O8RG54 saccharomyc
30	388	10.5	893	5	O76417	O76417 drosophila
31	387.5	10.4	179	2	O93OH1	O93OH1 streptococc
32	354	9.5	380	3	P78862	P78862 schizosacch
33	325.5	8.8	788	5	O9BLV2	O9BLV2 trypanosoma
34	280.5	7.6	338	11	O8RA07	O8RA07 mus musculu
35	278	7.5	695	3	O07980	O07980 saccharomyc
36	277.5	7.5	1151	10	O81785	O81785 arabidopsis
37	276.5	7.5	389	4	O16530	O16530 homo sapien
38	266.5	7.2	252	4	O15157	O15157 homo sapien
39	266.5	7.2	715	3	O12083	O12083 saccharomyc
40	265	7.1	186	4	O16603	O16603 homo sapien
41	261	7.0	389	4	O13400	O13400 homo sapien
42	254	6.8	297	4	O95744	O95744 homo sapien
43	253	6.3	630	5	O8SVO0	O8SVO0 encephalito
44	232.5	6.3	166	4	O96H10	O96H10 homo sapien
45	231	6.2	119	3	O96V63	O96V63 magnaporthe

ALIGNMENTS

RESULT 1

ID	PRELIMINARY:	PRT:	715 AA.
O8S2D3	O8S2D3		
AC	O8S2D3:		
DT	01-JUN-2002 (TREMBLrel. 21, Created)		
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Putative MHL1.		
GN	P0401G10.8.		
OS	Oryza sativa (japonica cultivar-group).		
OC	Eukaryota; Viridiplantae; Magnoliophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
OC	Ehrhartoideae; Oryzeae; Oryza.		
OX	NCBI_TaxID=39947;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CV. NIPPONBARE;		
RA	Sasaki T., Matsumoto T., Yamamoto K.;		
RT	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC		
RT	clone: P0401G10.8.		
RL	Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF003238; BAB89000.1;		
SQ	SEQUENCE 715 AA; 79943 MW; 75F99169C66F41AC CRC64;		

Query Match	98.3%	Score 3645;	DB 10;	Length 715;
Best local similarity	99.9%	Pred. No. 1.8e-233;		
Matches 712; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;
OY	1	MDESPRGGCGACGPEPRIRLEESVVRNIAAGVIRPSSAVKELIENSIDACASSVYA	60	
DB	1	MDESPRGGCGACGPEPRIRLEESVVRNIAAGVIRPSSAVKELIENSIDACASSVYA	60	
OY	61	VNDGKGLIOVSDGGRIFEDIALICERTTSKLSAYEDLQTKSMGFEGLASMTYV	120	
DB	61	VNDGKGLIOVSDGGRIFEDIALICERTTSKLSAYEDLQTKSMGFEGLASMTYV	120	
OY	121	GHTVTTTTEGOLHGVRSVROGVMEPRCAAVGCTGVNENLYNNVAKKTKLQNSN	180	
DB	121	GHTVTTTTEGOLHGVRSVROGVMEPRCAAVGCTGVNENLYNNVAKKTKLQNSN	180	
OY	121	GHTVTTTTEGOLHGVRSVROGVMEPRCAAVGCTGVNENLYNNVAKKTKLQNSN	180	
DB	121	GHTVTTTTEGOLHGVRSVROGVMEPRCAAVGCTGVNENLYNNVAKKTKLQNSN	180	

SEQUENCE FROM N.A.
 RA Horvath M., Hilpert M., Kunze R.;
 RT "Atmhl", an Arabidopsis homologue of the E. coli mutL DNA mismatch
 RT repair gene."
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vil D.M.,
 RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
 RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ012747; CAI0163.1; -
 DR EMBL: AL161514; CAB78038.1; -
 DR HSSP: P23367; 1BKN
 DR InterPro: IPR003594; ATPbind_Atpase.
 DR InterPro: IPR002099; DNA_mis_repair.
 DR Pfam: PF01119; DNA_mis_repair.1.
 DR Pfam: PF02518; HATPase_c.1.
 DR TIGRFAms: TIGR00585; mutL.1.
 DR PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; UNKNOWN_1.
 DR NON_TER 1
 SQ SEQUENCE 737 AA; 82364 MW; 0497793964CA3700 CRC64;
 Query Match 67.5%; Score 2505; DB 10; Length 737;
 Best Local Similarity 66.4%; Pred. No. 8,7e-158;
 Matches 482; Conservative 111; Mismatches 129; Indels 4; Gaps 3;
 QY 2 DEPSRGGGCGEPPRIIRLEESVYVNRVIAAGEVYQRPSSAVKELIENSLSAGASSVYV 61
 DB 13 EEESPATVIVREBPVKIQRLEESVYVNRVIAAGEVYQRPSSAVKELIENSLSAGASSVYV 72
 QY 62 KDGLKLIQVSDGHRPREDALICRHTTSKISATEDLOTISMGFRGALASMTYVG 121
 DB 73 KDGLKLIQVSDGHRPREDALICRHTTSKISATEDLOTISMGFRGALASMTYVA 132
 QY 122 HWYVTTTECOLHGYRYSYRDGVNENPKCAAYKGTQVWENLFYMWVARKKTLQNSND 181
 DB 133 HWYVTTTQKQIHGYRYSYRDGVNENPKCAAYKGTQVWENLFYMWVARKKTLQNSAD 192
 QY 182 DYKTIQVDSRFAVHNINVTFSCKRHGARNADYHASTSSRLDAIRSYGASVYRDLEI 241
 DB 193 DYKTIQVDSRFAVHNINVTFSCKRHGARNADYHASTSSRLDAIRSYGASVYRDLEI 252
 QY 242 KVSIEDAADSTFKMDGYSINANYAKITMLFINDRLVDCSTALRAIEFYSATLPQAS 301
 DB 253 EVSSCDSSGCTFDEMEGFIINSNYAKKITLVLFINDRLVECSALRAIEFYAATLPKAS 312
 QY 302 KPIFYMSIHLPEHADVNIHPTKEVSLNQERITETIRNAIEEKIMNSNTRTFOTQAL 361
 DB 313 KPIFYMSIHLPEHADVNIHPTKEVSLNQERITETIRNAIEEKIMNSNTRTFOTQAL 372
 QY 362 NLSGIAANPOKDKVSEASMGSGTKSOKIPYSQWVRTPRNPSCRLLHYHGGSSNLEKK 421
 DB 373 EYIIO-SLTLSQKSDSPVSOGRSGQTKQVPAKAVRTDSSDPAGRLAFLQPKQSLPDK 431
 QY 422 FDLVS--RNVYVRSRRKQKADGDLSSRHELLYEIDSSFFHGLDLYVKNCTYGLADEAFAL 480
 DB 432 VSSLSVYRVSRRKQKADGDLSSRHELLYEIDSSFFHGLDLYVKNCTYGLADEAFAL 491
 QY 481 IOHNTRLYLNVVNIISKELMYOALCFRGENATQISEPAPLQELVWALDKDEL--MSD 538
 DB 492 VOYFTHTLYIANVVALSKELMYOALCFRGENATQISEPAPLQELVWALDKDEL--MSD 551
 QY 539 EKDEKLEIAEVNTIELKENAMINERYSHIHODGKITRLPVVLDQYTPDMRLPEFVL 598
 DB 552 TKDCLKREIAEMNTIELKENAMINERYSHIHODGKITRLPVVLDQYTPDMRLPEFVL 611
 QY 599 ALGNDVYVWDEKECFRTVYASAVGNFYALHPILNPSGNGHILKKKNDSDAHDHAEIDL 658

RESULT 4
 Q8VDI4 PRELIMINARY; PRT; 760 AA.
 ID Q8VDI4
 AC Q8VDI4
 DT 01-MAR-2002 (TREMUREL. 20, Created)
 DT 01-MAR-2002 (TREMUREL. 20, Last sequence update)
 DT 01-JUN-2002 (TREMUREL. 21, Last annotation update)
 DE Hypothetical 84.7 kDa protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC021815; AAH21815.1; -
 DR InterPro: IPR003594; ATPbind_Atpase.
 DR InterPro: IPR002099; DNA_mis_repair.
 DR Pfam: PF01119; DNA_mis_repair.1.
 DR Pfam: PF02518; HATPase_c.1.
 DR TIGRFAms: TIGR00585; mutL.1.
 DR PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; UNKNOWN_1.
 KM Hypothetical protein.
 SQ SEQUENCE 760 AA; 84693 MW; AAA69BD99E0F3B6 CRC64;
 Query Match 38.8%; Score 1438; DB 11; Length 760;
 Best Local Similarity 38.5%; Pred. No. 5,9e-87;
 Matches 303; Conservative 153; Mismatches 218; Indels 112; Gaps 12;
 QY 18 IRLEESVYVNRVIAAGEVYQRPSSAVKELIENSLSAGASSVYVAKDGKLIQVSDGHR 77
 DB 8 IRLEDETVVNRVIAAGEVYQRPSSAVKELIENSLSAGASSVYVAKDGKLIQVSDGHR 67
 QY 78 IRFEDLALICRHTTSKISATEDLOTISMGFRGALASMTYVGHVYVTTTECOLHGYR 137
 DB 68 IRFEDLALICRHTTSKISATEDLOTISMGFRGALASMTYVGHVYVTTTECOLHGYR 127
 QY 138 VSYRDGVNENPKCAAYKGTQVWENLFYMWVARKKTLQNSNDYKPIVDISRFVAVH 197
 DB 128 ASYDGLQAPPPKCAAYKGTQVWENLFYMWVARKKTLQNSNDYKPIVDISRFVAVH 187
 QY 198 INVTFSCRHGANADYHASTSSRLDAIRSYGASVYRDLEIKVSEDAADSTFKMDG 257
 DB 188 SGISISYKQGETVSDVPTLPNATTVDIRSIFGNVARSRELEIYGCEDCKTALA--FKMNG 244
 QY 258 YISNANYAKKITMLFINDRLVDCSTALRAIEFYSATLPQASAPFYMSIHLPESEH 317
 DB 245 YISNANYAKKITMLFINDRLVDCSTALRAIEFYSATLPQASAPFYMSIHLPESEH 304
 QY 318 VNIHPTKEVSLNQERITETIRNAIEEKIMNSNTRTFOTQALNSGIA-----QANPQ 372
 DB 305 VNIHPTKEVSLNQERITETIRNAIEEKIMNSNTRTFOTQALNSGIA-----QANPQ 363
 QY 373 KDKVSEASMGSGTK-----SOKT-----PVSQWVRTPRNPSCRLLHYHGG-- 413
 DB 364 TGVASSSTSGSGDGVYAVQWVRTSDRQKIDAFLOPVSSILVSPQDPAPARGARTESGP 423
 QY 414 -----OSSNLEKKFVYVSRVNRV-----SRNOKDQGD 442
 DB 424 ERATREDEEMALAPAPAAEAASENLERE-SLMEISDTAQAAPFTSSPGSSRKRHRREDSD 482

[illegible]

DB 653 NLEPTLYKVERC 664

RESULT 6

061917 PRELIMINARY; PRT; 663 AA.
AC 061917.
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE MUTL homolog
GN MLH1 OR CG11482.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON S;
RA McKee B.D., Borromeo D.;
RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF068257; AAC19117.1; -
DR HSSP: P23367; 18KN.
DR Flybase: FBgn0011659; Mhl1.
DR InterPro: IPR003594; ATPbind_ATPase.
DR InterPro: IPR002099; DNA_mis_repair.
DR Pfam: PF01119; DNA_mis_repair; 1.
DR Pfam: PF02518; HATPase_C; 1.
DR TIGRfams: TIGR00585; mult; 1.
DR PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; UNKNOWN_1.
SQ SEQUENCE 663 AA; 75848 MW; 590A666B2DB9D80B CRC64;

Query Match 33.9%; Score 125.5; DB 5; Length 663;
Best Local Similarity 36.7%; Pred. No. 6,2e-75;
Matches 269; Conservative 145; Mismatches 223; Indels 95; Gaps 11;

QY 14 EPPRIRLESYVNRITAGEVIOEPSAVKELIENSLDAGASSVAVKDGKILQVSD 73
DB 6 IGVVIRKLDDEVVNRITAGEVIOEPSAVKELIENSLDAGASSVAVKDGKILQVSD 65
QY 74 DGHGIFEDLALICERHTTSKLSAYEDLOTISMGREGALASMTYGVHTVTTTEGOL 133
DB 66 NGGIGIRDELATVYCERFTSKLRFEDLSQIAFFGREGALASISVHMLSTQTKAKEK 125
QY 134 HGTVSYRDGVNENRKPAAVKGTOVWENLFYNNVARKKTIONSNDYPKIVDISRP 193
DB 126 CGKATYADGKLOGKPCAGNOGITICIEDLFYNNPQRORLSPAEEFORLSEVLAKY 185
QY 194 AVHHINVTSCRRKHGHRADVHSASTSSRLDAIRSVYAGSVVDLIEIKVSYEDADSI 253
DB 186 AVHNPRVGFTRKQGOAOPALRTPVASSSENRIITYGAISKELEF-----SHNDEV 240
QY 254 KMDG--YISNANVAKITMLFINDRLVDCATLAKRAIEFVSATLPQASKPTIYMSIHL 311
DB 241 KEFAECILITOVNSAKCMLELFINOLVESTALRTSVDSIYATYLPGRHHFVYMSLTL 300
QY 312 PSEHVDVNIHPTKEVSLNNOERIETIRNAIEEKLMSNTTTRIFOTQALNSGIQANP 371
DB 301 PPNLDVNIHPTKEVSLNNOERIETIRNAIEEKLMSNTTTRIFOTQALNSGIQANP 371
QY 372 QKDVSEASGSGTKSOKIPVSQWRTDP-----RNPGRHLYWHGSSNL 418
DB 356 DDEDTQAD-----KTQRIYPRKMYRTDSTEQKLDKFLAPLVKSDSGVSSSQEASRLP 410
QY 419 EKKFDLVSVANVRNRNOKDAGDLSRRHLEVIDSPHGLDILVKNCTYVGLADAP 478
DB 411 EESFRTYAAK-----KSREVR-----LSVLDMRKRVKQSVQRLTKNLVYGVCDERR 462
QY 479 ALIQHTRLIYLVNVIKELMQALCRGNNAIQLSPAPLOELVWALDDELMSD 538
DB 463 ALFOHETRLYMCNTRSFSEELFYQMIYEFONCEITICPILPKELLILLESRAAGWT 522

QY 539 EKDEKLEIAEVNTEILKENAEMINEYFSIHIDQDKLTRLPVVLDOQYPPMDRLEFVYL 598
DB 523 PEDEKAEILADGAADILKKAPIMREYFGLRISDEQDMSLSEFLHQHRCVAAHLEFVYL 582
QY 599 ALGNDVTWDEKRCERTVASAVGNFYALHPILPNFSGCINHLYKKNRDSMADEHENDI 658
DB 583 RLATEVDWDEKRCERTVETPRETARY-----
QY 659 ISDENVDDELLAEADAANAQ-----REWTIOHVLFPMSRLFKPKMSMATDGTFOVA 712
DB 609 -----AQLDMREGATAVFSRWTMEHVLPPAKKTLPPR---IKQDIYELT 651
QY 713 SLEKLYKIFERC 724
DB 652 NLEPTLYKVERC 663

RESULT 7

09P7W6 PRELIMINARY; PRT; 684 AA.
AC 09P7W6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative DNA mismatch repair protein, Mhl1 homolog.
GN SPBC1703.04.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA McDougall R.C., Rajandream M.A., Barrell B.G., Cadieu E., Lelaure V.,
RA Gallibert F.;
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AL136536; CAB66448.1; -
DR HSSP: P23367; 18KN.
DR InterPro: IPR003594; ATPbind_ATPase.
DR InterPro: IPR002099; DNA_mis_repair.
DR Pfam: PF01119; DNA_mis_repair; 1.
DR Pfam: PF02518; HATPase_C; 1.
DR TIGRfams: TIGR00585; mult; 1.
DR PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; UNKNOWN_1.
SQ SEQUENCE 684 AA; 77253 MW; 7A59633B36E2944B CRC64;

Query Match 28.1%; Score 1040.5; DB 3; Length 684;
Best Local Similarity 32.7%; Pred. No. 1,2e-60;
Matches 241; Conservative 149; Mismatches 259; Indels 87; Gaps 13;

QY 17 RIRRLSESVNRITAGEVIOEPSAVKELIENSLDAGASSVAVKDGKILQVSDDG 76
DB 8 KIRPLDQVLINKIAAGEIIRPENAIKELIENSLDAGSSIDVLKDGKILQVSDH 67
QY 77 GIFEEDLALICERHTTSKLSAYEDLOTISMGREGALASMTYGVHTVTTTEGOLHGY 136
DB 68 GIQYDLPYLCKRFSTSKINDNDLOHLOTFTGREGALASISVHAKVTYTKLSSDTHAW 127
QY 137 RVSYRQGV-----MENEPKPCAAGVKGTOVWENLFYNNVARKKTIONSNDYPKIVDP 189
DB 128 KATYVGGALAPISPGMSAPQPCAGNOGITITAEDELFYNNRKRKSAKNSSEFRRTMIL 187
QY 190 ISRPAVHHINVTSCRRKHGHRADVHSASTSSRLDAIRSVYAGSVVDLIEIKVSYEDAA 249
DB 188 VQRYAIHNDQVSPNCCKKVGDTVASLSSRLSKADKIRHRYGPRVASHLDFLGGQSS 247
QY 250 DSIEFKMDGYSNANVAKITMLFINDRLVDCATLAKRAIEFVSATLPQASKPTIYMSI 309
DB 248 IYGFSAAGFISNADPODKSNLILFTINRLVESVELHHALEETYAKYLHNGASYFYLSL 307
QY 310 HPSHVDVNIHPTKEVSLNNOERIETIRNAIEEKLMSNTTTRIFOTQAL--NLGIA 367

[illegible]

RL Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode *C.elegans*: A platform for
 investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL; 292813; CAB07283.1; -
 DR HSSP; P23367; 1BN.
 DR InterPro: IPR003594; Arpbind_Artpase.
 DR InterPro: IPR002099; DNA_mis_repair.
 DR Pfam: PF01119; DNA_mis_repair; 1.
 DR Pfam: PF02518; HATPase_C; 1.
 DR TIGRfams; TIGR00585; multi; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; UNKNOWN.1.
 SQ SEQUENCE 779 AA; 88104 MW; 05038F5253306CA4 CRC64;

 Query Match 22.8%; Score 845; DB 5; Length 779;
 Best Local Similarity 29.7%; Pred. No. 1.4e-47;
 Matches 240; Conservative 136; Mismatches 276; Indels 156; Gaps 22;

 QY 18 IRRLESVNRIAGEVIRPSSAKELIENSIDAGASSVAVKDGKLIQVSDGHC 77
 DB 25 IORLPDQVNRMAAGEVIRPCKAIKELVENSIDAGATEIMVMONGKLLQVSDNGK 84
 QY 78 IFFEDALICERHTTSKLSAYEDLIOTIKSMGFEGEALASMTYGVYVTTTTEGQLHGYR 137
 DB 85 IREDEPVALCERNATSKLOKPEELMHMKTYGFGELALSLSHAKNIYISKRADACANQ 144
 QY 138 VSTRGVMEKNEPCAAVGVTOVMEVNTFYNNYARKKTLQNSNDYPKIVDFISRAVHH 197
 DB 145 ANFLDCKMADTKPAAGKGTCTATDLEFNLPTRRNKMTTHGEAKMNDTLREPAIRH 204
 QY 198 INTFSCRHGANRADVHASTSSRLDAIRSVYGASVVDLEIKYSIEDAASIRKMG 237
 DB 205 PVSFSLRQNA--GDFRKGGDNFQVYVONLGRVADTLTSL--NSTRLKTFYTG 239
 QY 258 YTSN-----ANYVAKITMILFINDRLVDCALRAIEFYVSAFLPQASKEPIYM 307
 DB 260 HSKRPASATAIAQNRKTSRSFVSFINGRSVRCOLIKHPIDEVLCAR--QLHAQFCL 317
 QY 308 SIHPSBEADVNIHPTKREYSLNORITETITNAIEELKMSNTRITRITQOLANLSCGA 367
 DB 318 HQIDETRIDVNVHPKNSVITFLEKEIIEIR-AFFEYVIG---EIRGFALDVEKE 372
 QY 368 QANP-----OKDVSEA-----SMGSGKSKOKIPVSGM-VETDR--- 401
 DB 373 EEOPIEENLWIPMSQSLKSIKIRPDKPEFKSSPSAMKSKKRVYDMEVETDAKERK 432
 QY 402 -----NPSGRLL-----HTYMHGSSNLKKEKFDLVSVRVV-- 431
 DB 433 IDEFVTRGAGVPTTNDLIFGSSGILKRAKRTDSTGGEKEPEDLWTDVDDVSMVLSV 492
 QY 432 ---RSRRNOKDAGD-----LSSRHLELVEIDSSFFHGLLDIYKNCYVG 472
 DB 493 ADCRRLNESODLGEDDDVDPEYKTHREFHESIEVLKREILANSQSILREMKTSFPG 552
 QY 473 LADDEALILQHNRLYLAVNVNISKELMQALCRGFRNALDLSPPA---LOELIYV 538
 DB 553 SINVKQLQFGTSLVHDEFTYLRFFQIYSVSGFNGSYRLDEPPIILIELEL 610
 QY 529 ALKDELMSDEKDEKLEI-AEYNT-----ELIKENAMENIYESFHIHO--DGKL--TR 578
 DB 611 ---GELSTREPRVYAFEFYFANYENRPAEKLAELHADLHDFALKLQDLNGRLHTE 666
 QY 579 LPVVLIDQYTPDMDRLEFEVYALAGNDYTWDEKECFRTVASAVNGFYALHPPILPNSGNG 638
 DB 667 IPELVYVFPQLEKLEPFLATVLANVDYDEQNTFRICRAIGDLFTL 714
 QY 639 IHIYKKNRSMADENHENDLSDEN--DVDQELLAEAMAMQREWTIOHVLFPSSRL 696
 DB 715 -----DTNFTITLDRKISAFSAIPW---KTLIREVLMLPVKKRF 749

QY 697 KPPKSMATDGTGFQVAVASLEKLYKIFERC 724
 DB 750 IPPEHFKQAGVIRQLADSHDLYKVFERC 777

 RESULT 10
 ID Q8TON1 PRELIMINARY; PRT; 504 AA.
 AC Q8TON1;
 DT 01-JUN-2002 (TRENBLrel. 21, Created)
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE GH1871/P.
 GN MLH1.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Staphenon M., Brokslein P., Hong L., Agdayant A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.,
 RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AT069160; AAL39305.1; -
 SQ SEQUENCE 504 AA; 58261 MW; 9F6CF36F70066FC CRC64;

 Query Match 21.2%; Score 788; DB 5; Length 504;
 Best Local Similarity 31.4%; Pred. No. 4.4e-44;
 Matches 181; Conservative 111; Mismatches 191; Indels 94; Gaps 11;

 QY 169 MVARKTLQNSNDYKIVDFISRAVHHINTFSCRHGANRADVHASTSSRLDAIRS 228
 DB 1 MPORRQALSPAEFEORLEVLARVAVHNPRVGFTRKQDAPALTRPVASRSRENI 60
 QY 229 VVGASVVRDLIEKYVEDADSIKMDG--YISNNYVAKKITMILFINDRLVDCATLK 286
 DB 61 ITGAASKELLEF-----SHRDEVYFEAECLTYQVNSAKCOMLFIQRLVSTALR 115
 QY 287 RAIEFYVATLPOASKPEIYSHLPSBEADVNIHPTKREYSLNORITETIRNMIIEK 346
 DB 116 TSVDSIYATILPGRHPFYMSLTPPQNDVNVHPKHEVHLVQBEIVDSIKQOYEAR 175
 QY 347 LKNSNTRITRITQOLANLSGIAQANPQKDVSEASMSGSTKSKRTIPYQAVRTDP----- 400
 DB 176 LGNSNTRITRITQOLANLSGIAQANPQKDVSEASMSGSTKSKRTIPYQAVRTDP----- 400
 QY 401 -----RNDGRLHYTHWGSSNLKKEKFDLVSVRVNRSRRNOKDAGDLSRHELLET 433
 DB 226 KFLAPLVKSDSGVSSSSQEPASRLPEESFRVTAKE--KSREYR-----LSSVLDMKRY 277
 QY 454 DSGFHPGLDIYVNCYVGLADAPALIQNNTLILYLVNVNISKELMQALCRGFRNFA 513
 DB 278 EROCSQVLNSTLNLVYGVCDERRALFQHEIRLYMCTRFSFEELFYQMIYEFONCSE 337
 QY 514 IOLSEPAILOELVMAKDELMSDEKDEKLEIAEYNTIELKENAMENIYESFHIHO 573
 DB 338 ITISPLPLKELILISLSSAAQWTFEDGKALADGADILAKKAPIMEYVGLRISD 397
 QY 574 GKILRLPVVLIDQYTPDMDRLEFEVYALAGNDYTWDEKECFRTVASAVNGFYALHPPILPN 633
 DB 398 GMLESPLSLHQHRCVANHLPVYLRLATEVDEOETRCFETCRETARY----- 448
 QY 634 PSNGNHLVKKKNDSMADEHENDLSDENDVQELLAEAMAMQREWTIOHVLFPSSRL 697
 DB 449 -----AQLDMREGATAGFSRWTHY 469
 QY 688 LFPSSRLFLKPPKSMATDGTGFQVAVASLEKLYKIFERC 724


```

DB 234 GFVAKPEHSRKNHYS-----IFINGRYIKNEMLNKALIEGYHTLITIGREPICTYIN 287
OY 310 HUPSEVVDVNIHPTKEVSLNORIEITIRNAIEEKLMSNTTRIFQOAL-----NLG 365
DB 288 EMPDILVDVNHPTKEVSLNORIEITIRNAIEEKLMSNTTRIFQOAL-----NLG 365
OY 366 IQANPOKDKVSAWSGSGTSSOKIPVSOVVRDPNPSGRRLTYHGGSSNLEKFFDL 425
DB 339 V-----PKKKVHLHS-----PEOOKIEEORONTE-----NNOKETFSSE 373
OY 426 SVANVVRNRNOKDAGDLSRRHLEVIDSFP 459
DB 374 ESNKRFEMENOND-----EIVAKED-SYMP 398

RESULT 13
OBS500
AC OBS500: PRELIMINARY: PRT: 563 AA.
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE DNA mismatch repair protein (MUTL/HEXB family).
GN EC005.0300.
OS Encephallitozoon cuniculi.
OC Eukaryota; Microsporidia; Unikaryoniadae; Encephallitozoon.
OX NCBI_TaxId=6035;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope:
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA MEDLINE=21576510; PubMed=11719806;
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA Prensier G., Barbe V., Peyretallade E., Brotlier P., Winkler P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Encephallitozoon cuniculi."
RT Nature 414:450-453(2001).
DR EMBL: AL590445; CADD2547.1;
SQ SEQUENCE 563 AA; 65296 MW; 266536004495B0F6 CRC64;

Query Match 14.9%; Score 553; DB 5; Length 563;
Best Local Similarity 24.2%; Pred. No. 2e-28;
Matches 174; Conservative 133; Mismatches 241; Indels 172; Gaps 22;

OY 18 IRRLEESVNRNIAAGEVIOIPSSAVKELIENSLDAGASSVAVKDGKLIQVSDGHG 77
DB 3 IKRLPDSVIRISAGEVIRPYNIKETIENSLDANSTHTTMEQDGLTL-IVEDDGS 61
OY 78 IREDAIICERTTSSKLSAYEDLOTIKSMGRFGELASKTYGHVTV-TTTEGOLHG 136
DB 62 IHESDELLCKOYCTSKLKEELFSLSSVGEFALSSISRCARIKVSKRREGEL-GY 120
OY 137 RVSIRGVVNEBPCAAVKGTOVMEENLFYNNVARKKTLQNSNDYPRKIVDSPPAVH 196
DB 121 EAVYRRTKEIT-IKVGMDGDIIVEIKNFYNNKVAKEHPSKREIREMMVLGMISVF 179
OY 197 HINVTFSCKRHGANRADVHASTSSRLDAIRSYGASVVRDLIEIKVSYEDAASIFKMD 256
DB 180 NSRISFEL-FYGEKIQLELPKSRVCGEDGYSN-----EDRVKRVG--MLNEIYK 228
OY 257 G--YISNNANYA-----KTIIMILFINDRLVDCALAKRAIEFVSATIPQASKP 304
DB 229 GKLLFSDSEYELVIFSTQOFLCKGMVLVFNGLRVLVSOEMKESLEKYKDLLPPQKPL 288
OY 305 IYMSIHLPEHVDVNIHPTKEVSLNORIEITIRNAIEEKLMSNTTRIFQOALNS 364
DB 289 IYIELVKSMDVNVHPSKREVLFSNESMIOKLCCKAERLSKLD----- 335

```

```

OY 365 GIAQANPOKDKVSAWSGSGTSSOKIPVSOVVRDPNPSGRRLTYHGGSSNLEKFFDL 424
DB 336 --YEOKPLK-PLPKDSFQSPFIKYSDPTSQSI-----AECELEK-- 371
OY 425 VSVRNVVRNRNOKDAGDLSRRHLEVIDSFPGLDITVKNCTYGLADEAFALIQH 484
DB 372 -----TERRREFRLESLSKRLRLTYVDVTV-----FRSLSYGVADRPRTIIVHG 417
OY 485 TRILYVNVNISKEMLVQOALCRFENFALQISEPAPLOELVWALKDDELMDEKDEK 544
DB 418 SLLNKTYVPLKEIKYOSLINDGNF-----EKK 447
OY 545 LEIAEVNTEILKENAEMINEFSLIHIDQGLTRLPVLDQYTPDMRLPEFYALGNDV 604
DB 448 RLVRLCKMDKTRALNDYFSLIEVEE-HYIGIPIL-----STICID-ABE-LMSGFEI 500
OY 605 TWDEKECFRTVASAVGNFYALHPILPNPSGNGIHLKKNRDSMADEHAEMLDISND 664
DB 501 RRSSEYETLKNIDVSTLY-----SGVEKSTKL-----FNI 532
OY 665 VDELLAEAAQAQREWTQIHLFSPMRFLKPRSMATDGFVQVYASLEKIKIFERC 724
DB 533 LKRRITGTARL-----ECFGLVYTLKELYRNERC 563

RESULT 14
OBS56
AC OBS56: PRELIMINARY: PRT: 643 AA.
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE DNA mismatch repair protein mult.
GN FN0462.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriia; Fusobacterium.
OX NCBI_TaxId=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RA MEDLINE=21866394; PubMed=11889109;
RA Bhatnagar V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhatnagar V., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,
RA Larsen N., D'Souza M., Malunas T., Pusch G., Haselkorn R.,
RA Fongstein M., Kyriades N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RT J. Bacteriol. 184:2005-2018(2002).
DR EMBL: AE010558; AAL94658.1;
SQ SEQUENCE 643 AA; 73742 MW; D7CE5D30EBD38770 CRC64;

Query Match 14.5%; Score 539.5; DB 16; Length 643;
Best Local Similarity 28.4%; Pred. No. 2e-27;
Matches 173; Conservative 109; Mismatches 211; Indels 117; Gaps 20;

OY 17 RIRLEESVNRNIAAGEVIOIPSSAVKELIENSLDAGASSVAVKDGKLIQVSDGH 76
DB 9 RIRLDESVNNAIAAGEVMEPTSMIKELIENSLDAGSKIKLEVWNG--RDISISDSGC 67
OY 77 GIFEDLAIICERTTSSKLSAYEDLOTIKSMGRFGELASMTYGHVTVTTTEGOLHG 136
DB 68 GMSKEDLLSIEHATSKITRKEDLFNIRIYGRFGELASSIASKMLSSRREDMONCT 127
OY 137 RVSIRGVVNEBPCAAVKGTOVMEENLFYNNVARKKTLQNSNDYPRKIVDSPPAVH 196
DB 128 QMNVLGKQVTV-LKDIOKNVGTQIEIKDLFPYNNPARKKFLKENTETLNIKIDFLRALA 186
OY 197 HINVTFSCKRHGANRADVHASTSSR--LDATSYGASVVRDLIEIKVSYEDAASIFK 254
DB 187 NPVAKFTLNIEGK-----ESIKTSNGIENVALIEIKGNLKNFSKSL----- 230

```

Mon Apr 7 09:24:05 2003

us-09-954-950-2.rpt

Db 255 MDGYSNAN-VYAKKIMLFIINDRLVDTALKAIEFVYSATLPQASKPFIYMSIHLPS 313
OY 221 -GYLGNNALFKANRDSIFPFINGRSYKSYVEAVIAAHTKMLKGVYFALLFLEVER 288
Db 314 EHVADNHPKREYSLNQEELIETIRNAIE-----EKIMSNTRIFQ 347
OY 289 SEIDVAVHPSKRYKVEAQAQNIIFDLIGELIENEFTEDEFIISPIBAEVEENTKNF 347
Db 358 TQALNLSGIAQANPKQKVS-----ASMSGTSOK-----IPYSQVWRDPNPSGRILH 409
OY 348 -LDINDFKDMDQDELQLSVYGKYSKRYKNIYKESFEDINK--KIMT 395
Db 410 YMHGSS--NL-EKFEDLVSNVYRSR-----NOKAGDLSRHLLAVE 452
OY 396 FGSAGTTESTIMLNIEKEDSKNIENFDSREINDKVKOTIFNOEDTG--RGKIFPD 451
Db 453 IDSEFHPGLDIYKNTY--VGLADEAFALIOHNTRLYVNVNISKELMQALCRFGN 510
OY 452 FTS-----LKNIDFKVIGVFTFLIVERNGLEIYDQHIHERILYKLEKOEYN 502
Db 511 FNAIOLSPAPLOELV-----MALKDDEIMS-----DEKDDKLEIAEYNT 552
OY 503 HSMK-----QNLVATREFELDPREKQALENIEFISFGFDIDDEKKNILRSIPT 555
Db 553 ELKENAEMI 562
OY 556 MNRDSYEMI 565

RESULT 15
Q8RA70 PRELIMINARY: PRT: 590 AA.

AC Q8RA70: 2002 (Trembl, 21, Created)
DT 01-JUN-2002 (Trembl, 21, Last sequence update)
DT 01-JUN-2002 (Trembl, 21, Last annotation update)
DE DNA mismatch repair enzyme (predicted AtPase).
GN MUTL OR TPE135.
OS Thermoaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoaerobacteriales; Thermoaerobacteriaceae; Thermoaerobacter.
OC NCBI_TaxID=119072;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=MB47 / JCM11007;
RC MEDLINE=21992816; PubMed=11997336;
RX Bao Q., Tian Y., Li W., Xu Z., Huang L., Dong X., Ma Y., Ling L.,
RA Chen Y., Xue Y., Xu Y., Lai X., Yang H.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome."
RL Genome Res. 12:689-700(2002).
DR EMBL: AE013095; AAM24580.1;
KW Complete proteome.
SQ SEQUENCE 590 AA; 67196 MW; ASD01FB21A8A585E CRC64;

Query Match 14.3%; Score 530; DB 16; Length 590.
Best local similarity 25.7%; Pred. No. 7.4e-27;
Matches 149; Conservative 123; Mismatches 231; Indels 76; Gaps 14;

OY 17 RIRLEESVYRIAGEVYIOPSSAVKELIENSIDGASVSVAVDGKLIQVSDG 76
Db 3 KIHLDKTYVNIAGEVVERPASIVKELVNSIDAGSKTIYELLGIGPIKYVDDGC 62
OY 77 GIRFEDALICERTHSKISAYEDLOTIKSMGFGEALASMTYGHVYTTIEGOLHGY 136
Db 63 GMEIIDAVALFERHATSKIRSDDELFTITLIGFREGELASIAASKVYLOTKRENETFG 122
OY 137 RVSRYDGMENEPKPCAAYKGTQVWENLFYVAVARKKTLQNSDDYPKIVDPSRAVH 196
Db 123 RLVEGCKILEKTR-COCOGISVEKVDFFMTPARRKLKRPSTEAMVYEVVTRICLS 181
OY 197 HINTFSCKHGANRADVHASTSSRLDAIRSYGASVVRDLIEIKVSEDAADSIIFKMD 256

Db 182 NPGISFYVVD--KVOYFTSGNGSIEVILRLQKREYHSALEFSEFEADLKAKAFATK 239
OY 257 GYISNANYAKKIMLFIINDRLVDTALKAIEFVYSATLPQASKPFIYMSIHLPS 316
Db 240 NFILVSN-----RNMGFYVAGRYKKTLSAAYDEAFKTYPSDRYGVLLYLEINPRT 295
OY 317 DVNIHPTKREYSLNQEELIETIRNAIEEKIMSNTRIFQALNLSGIAQANPKQK 342
Db 296 DVNIHPSKLEVKFDDRRIFESYKRTIRALRESNL-----IPEVKLEDK 342
OY 376 VSEASMSGTSOKIPYSQVWRDPNPSGRILHYYHGGSSMLEKFD-LVSVNRYRSR 378
Db 343 NEGGQIGEVK-LSLPFEV-----KEKTDGALFVREYNT 378
OY 435 RNOKDAGDLSRHLLAVEIDSPFHPGLDIYKNTYVGLADEAFALIOHNTRLYLVN 494
Db 379 -EKKIDNAPKHE-----SSDSERNVKRLSDIRIYGLTFYVIEKGVFIIDQHA 430
OY 495 ISKELMQALCRFGNFNAIOLSEPAIQLMALKDDEIMSDKDEKLEIAEYNT 554
Db 490 AHERILYKLEKOEYNVQSRQVTFP-----IVVELQPGD-----METVGGEREL 474
OY 555 LKENAEMINEFSIHIDQKTRLPVLDQYTPMDRL 593
Db 475 LYKLGVEFEFGNSV-----VIREVPVILGO--PEAKKL 507

Search completed: March 27, 2003, 15:45:06
Job time: 47 secs